

1/27

## Features of a donor vector.

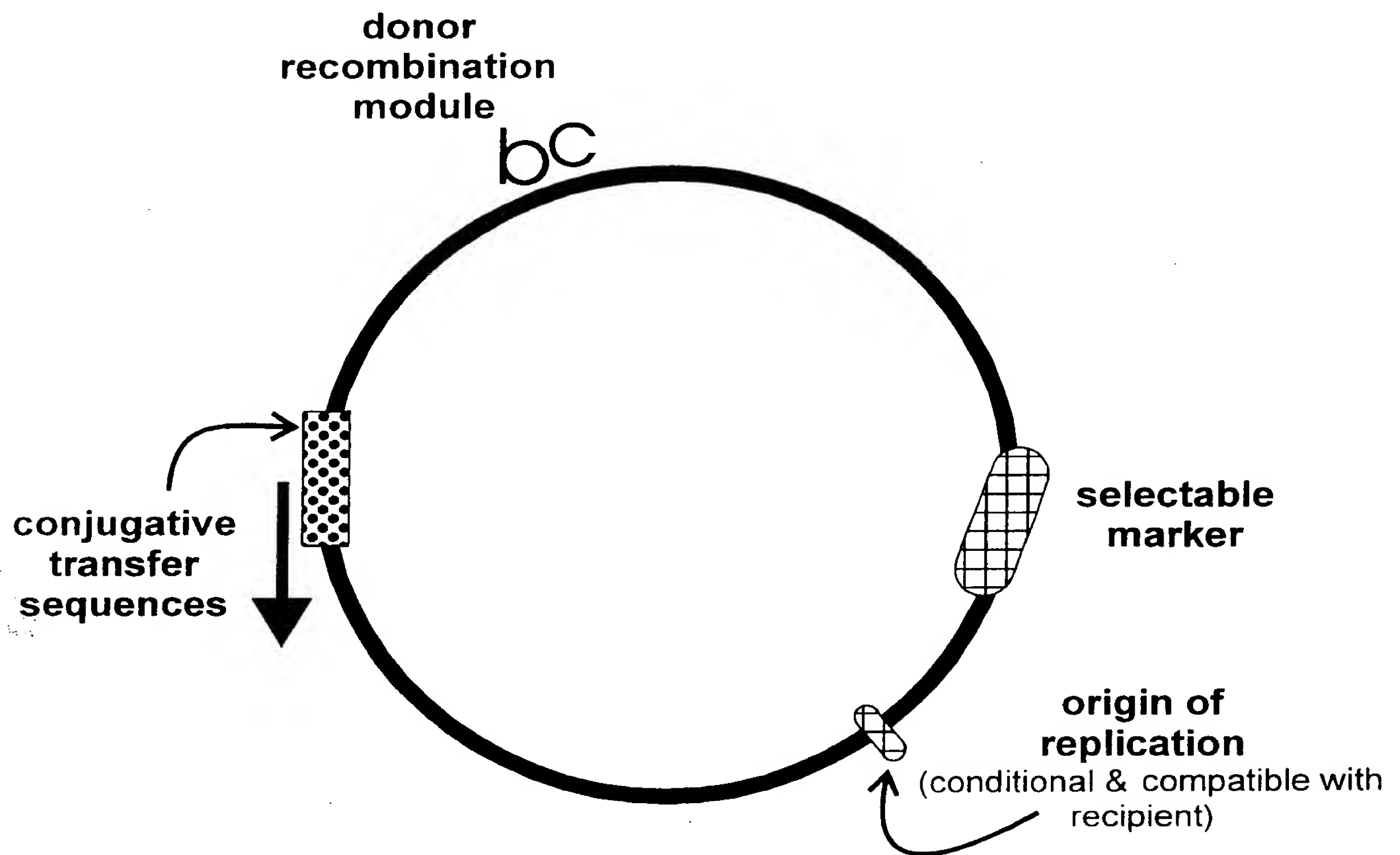
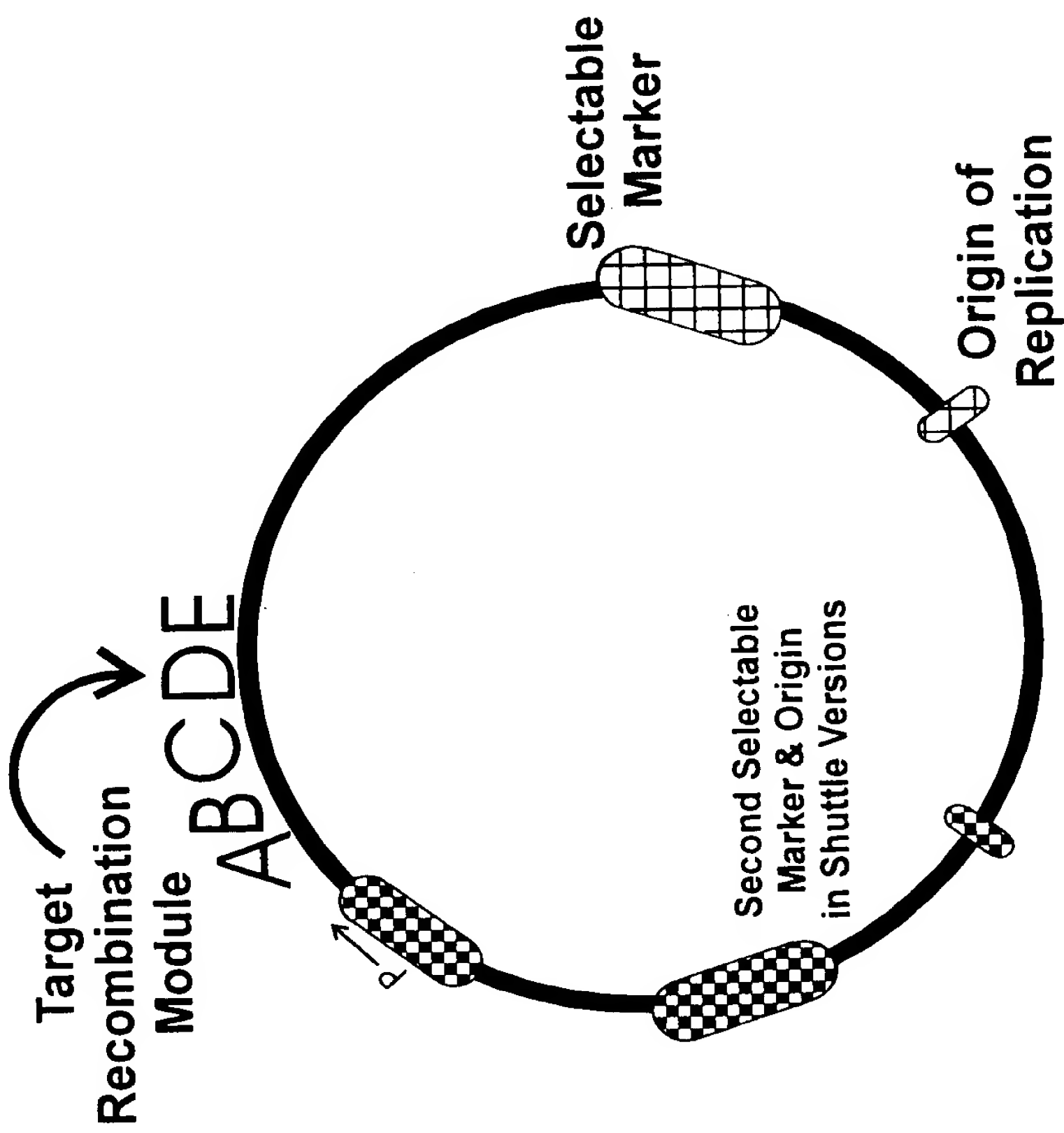


FIG.1

# Features of a target vector.



## Additional Features of DGA

GalK expression is lethal on galactose in a galE background. The insert is negatively-selectable.

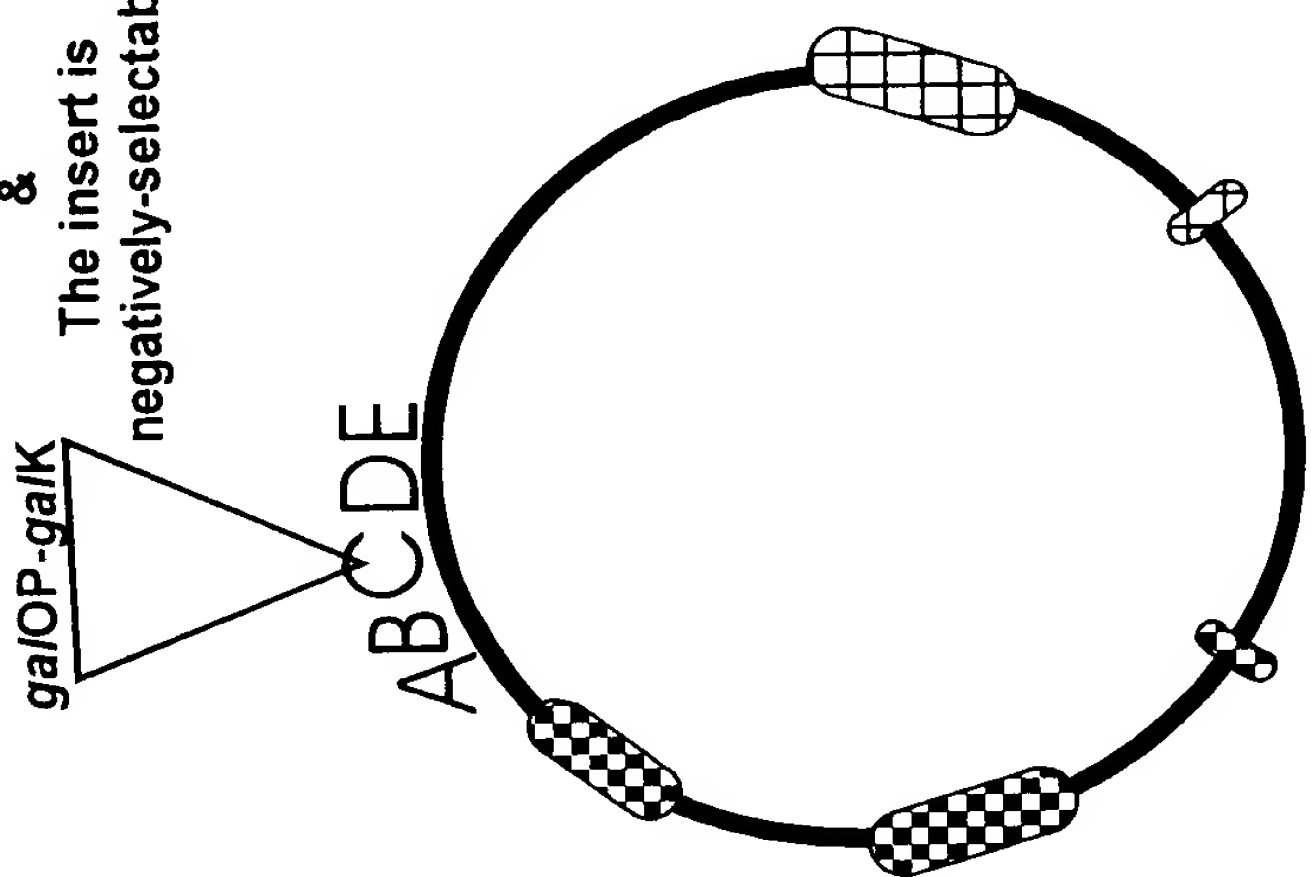
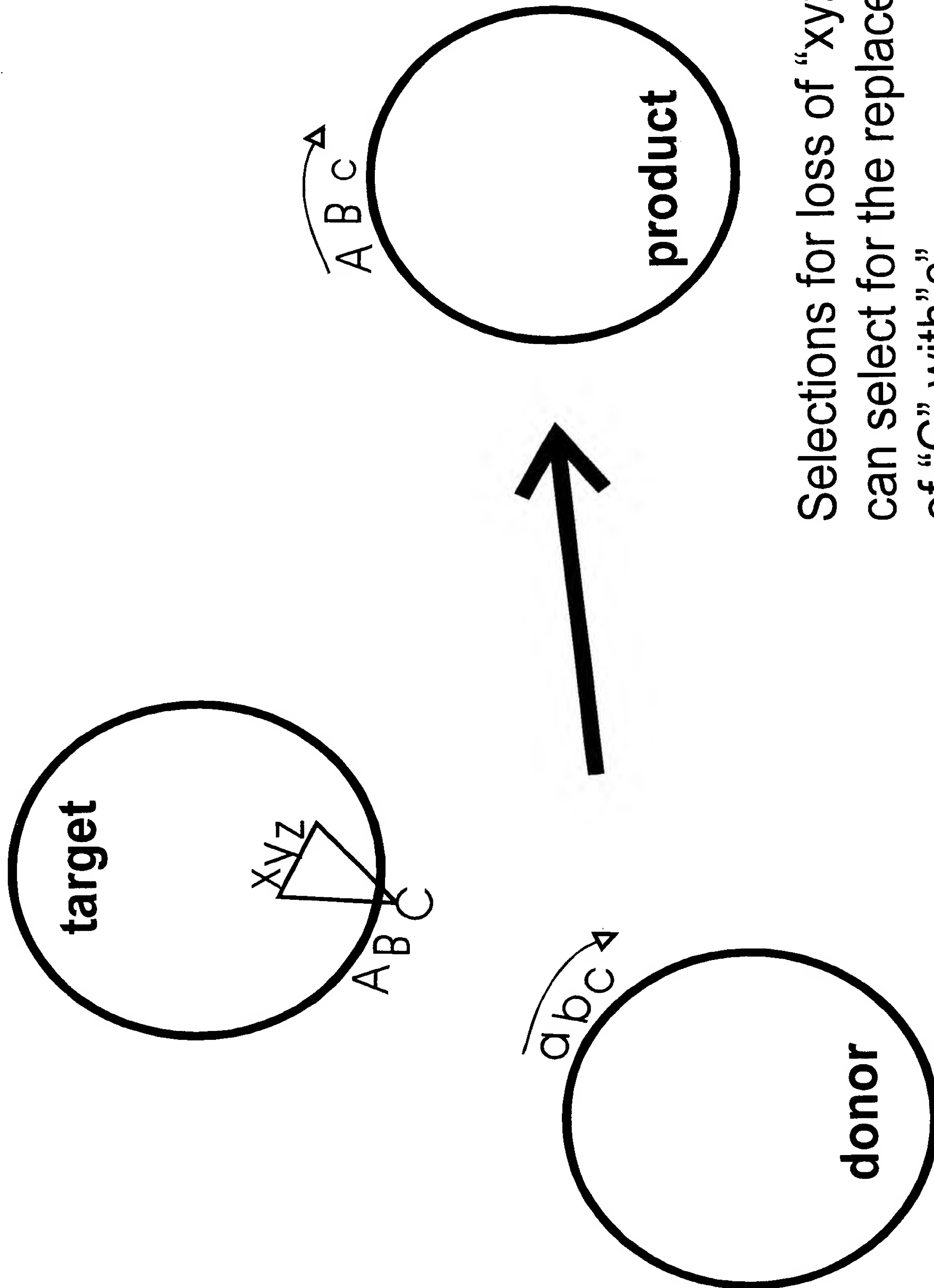


FIG.2

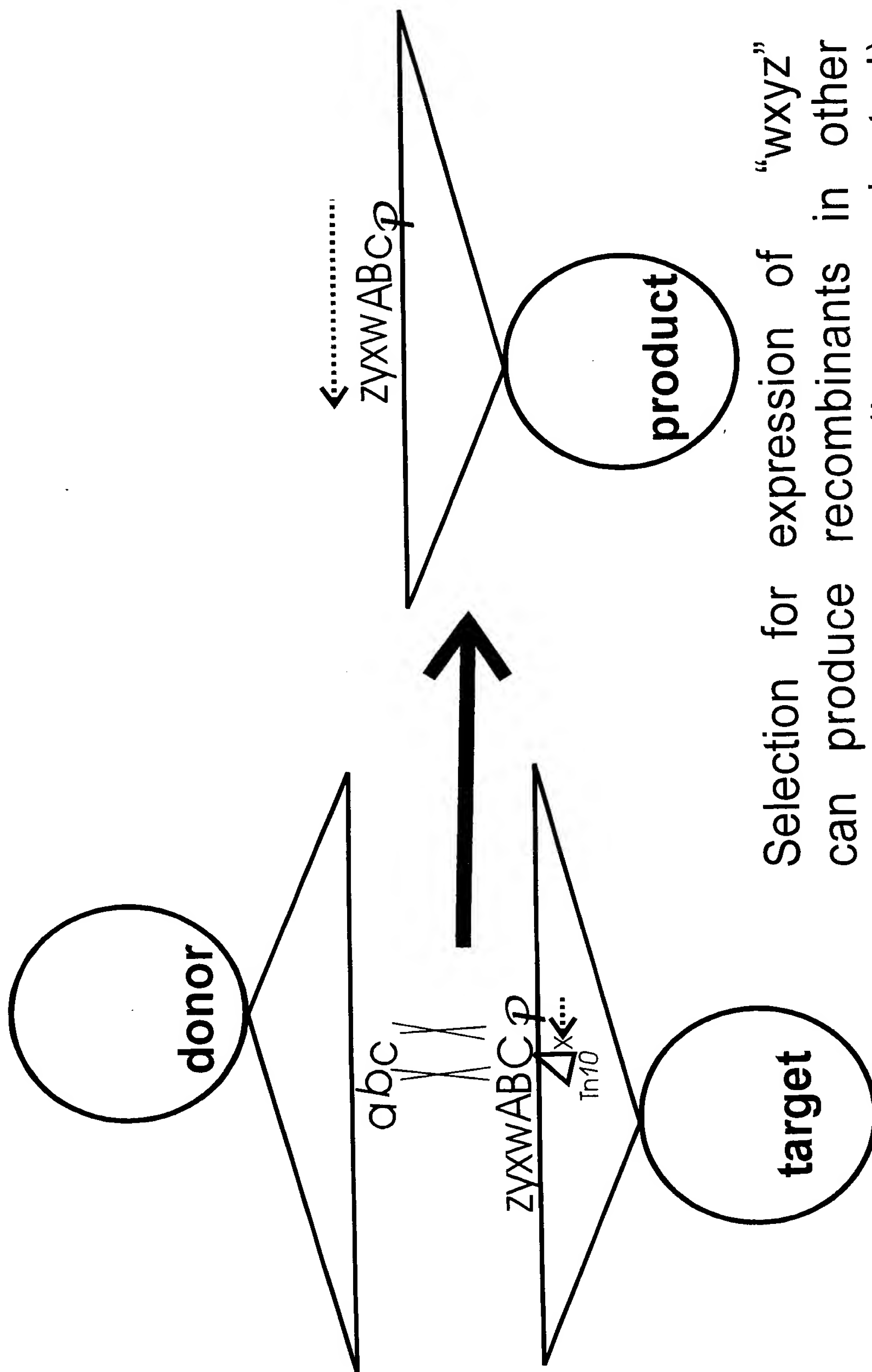
3/27



Selections for loss of "xyz"  
can select for the replacement  
of "C" with "c".

FIG.3

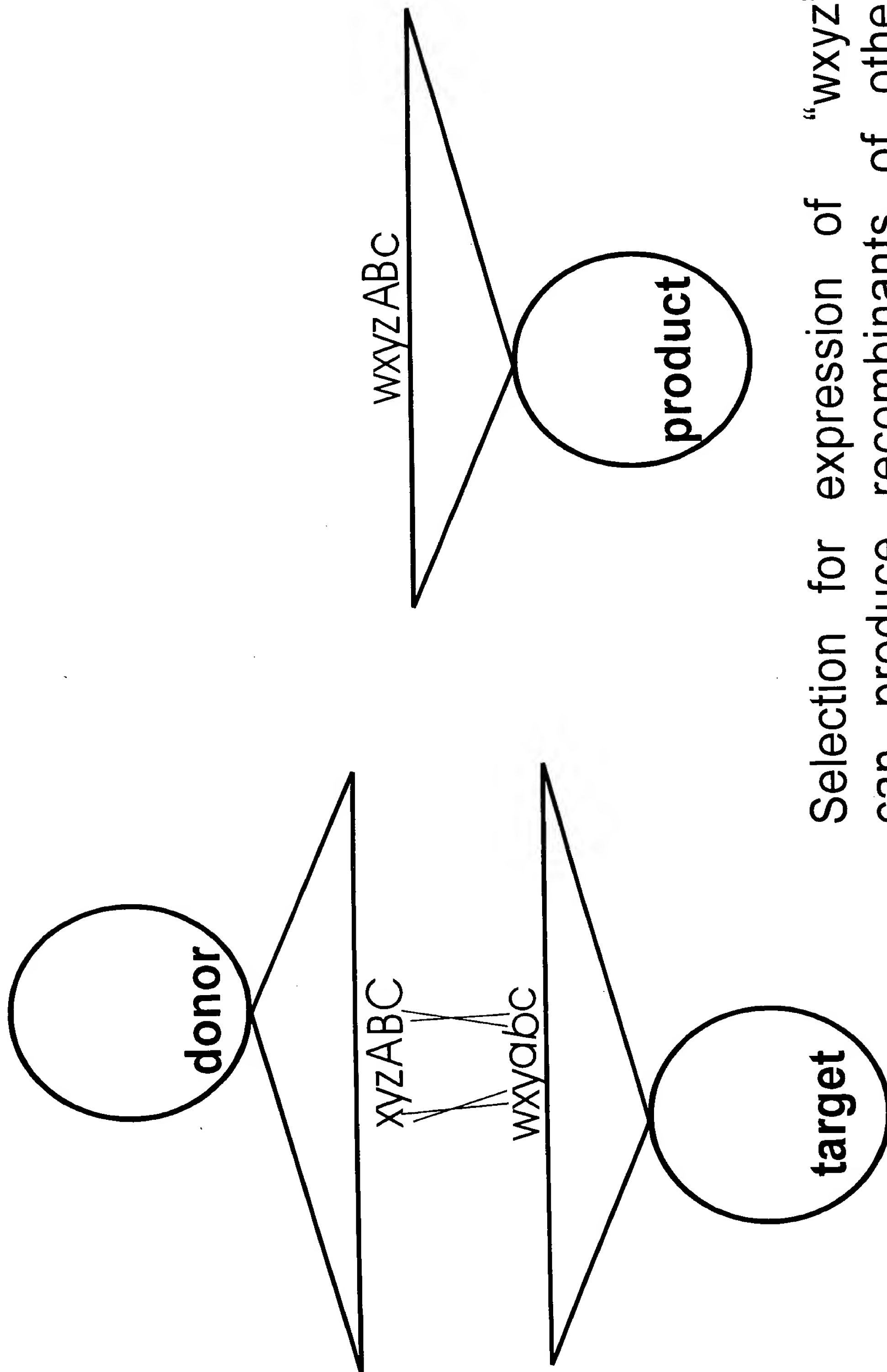
4/27



Selection for expression of "wxyz" can produce recombinants in other physically linked (but unselected) homologous sequences, such as, "ABc", in this example through removal of the polar Tn10 sequences.

FIG.4

5/27



Selection for expression of "wxyz" can produce recombinants of other physically linked homologous sequences, such as, "Abc" in the illustration.

FIG. 5

6/27

# Directed Gene Assembly applied to mutagenesis

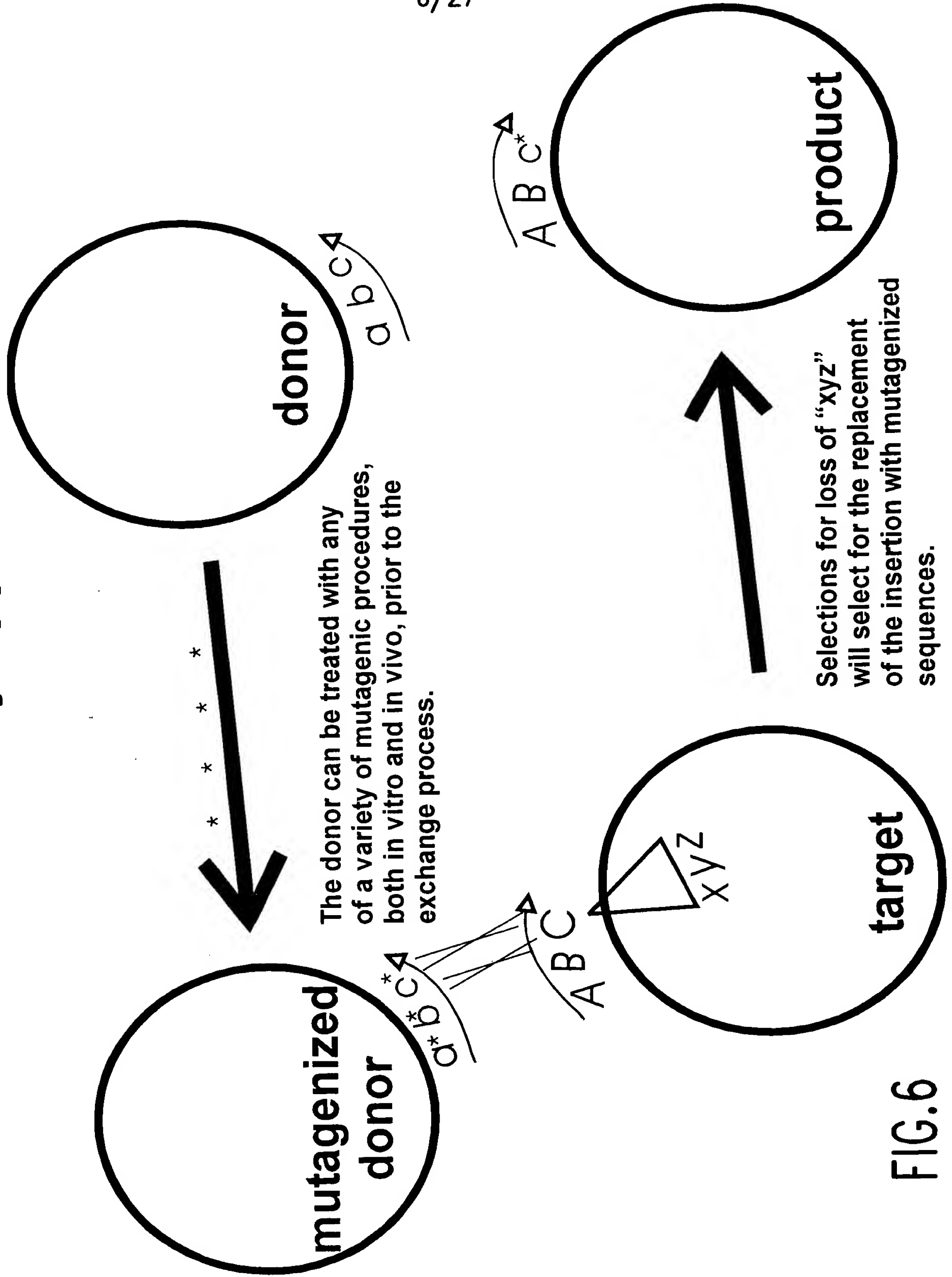
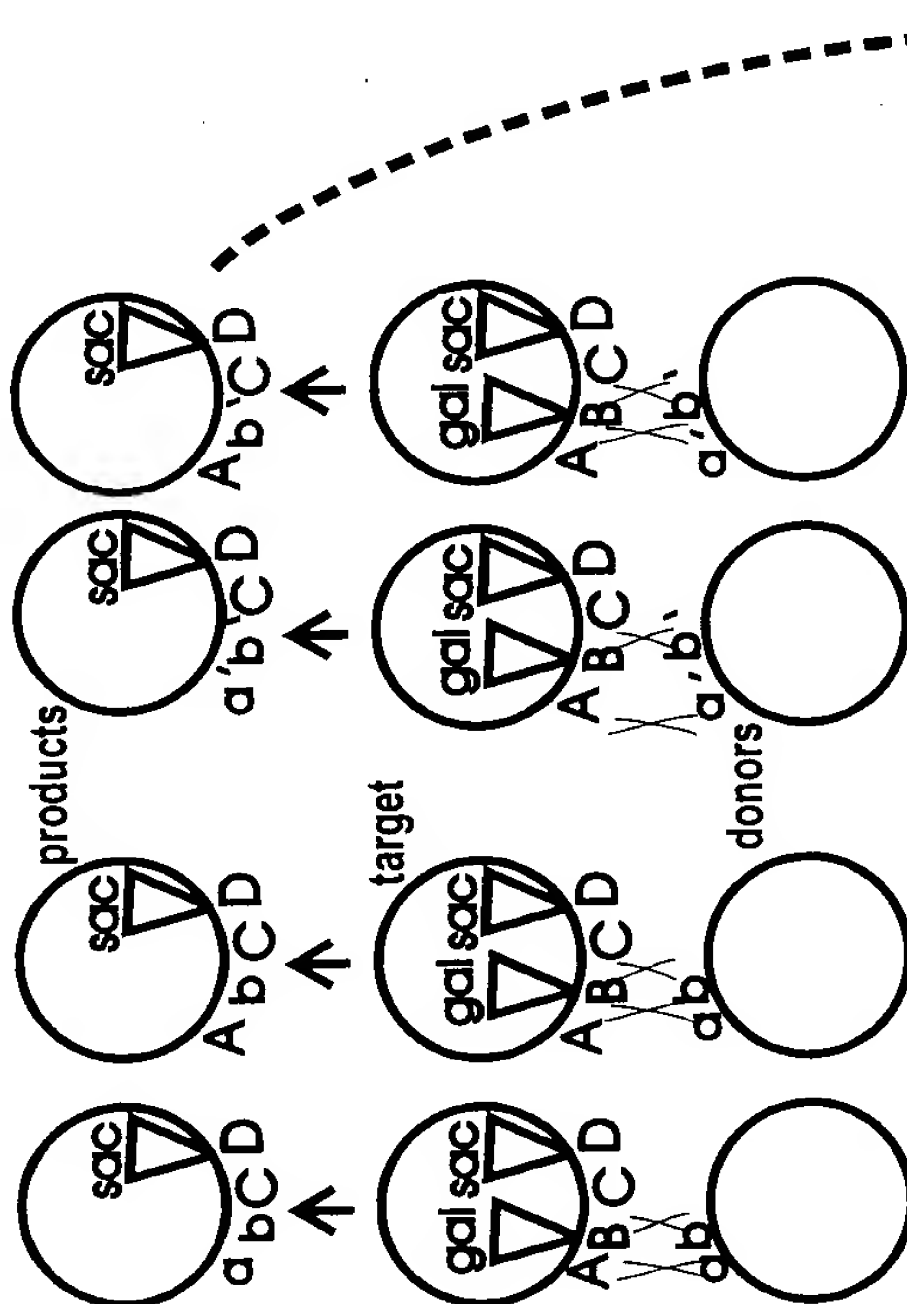


FIG.6

7/27

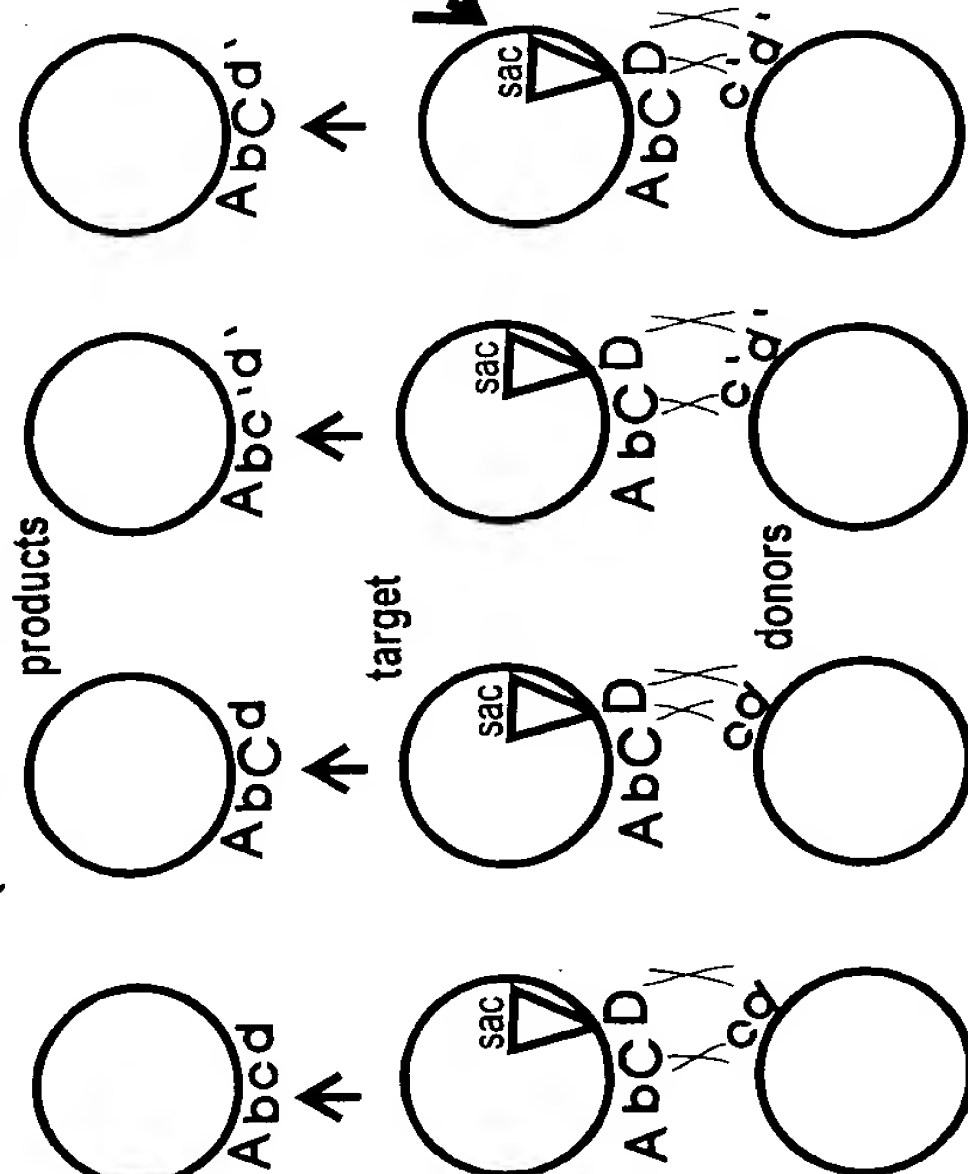
## First Product Series (galactose resistant)

Products of the first galactose resistant selected reaction still contain the sucrose insert disrupting the target.



## Second Product Series (sucrose resistant)

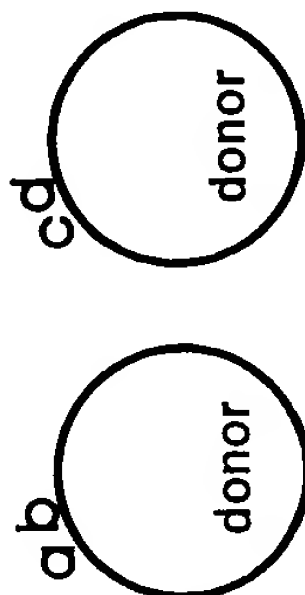
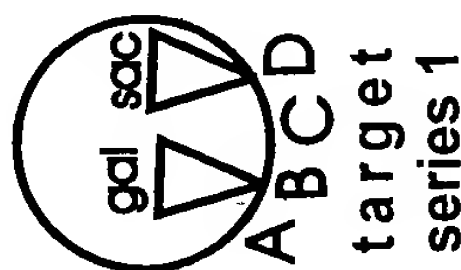
After the second sucrose resistance selection intact genes are reconstructed.



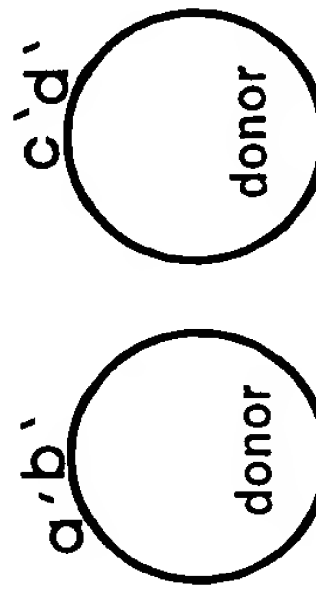
Products of the first reaction become the substrates of the second. Reactions with "AbCD" product are shown.

## Substrates

Target gene with family member #1 (complete) with two recombination-targeting inserts in targeting vector.



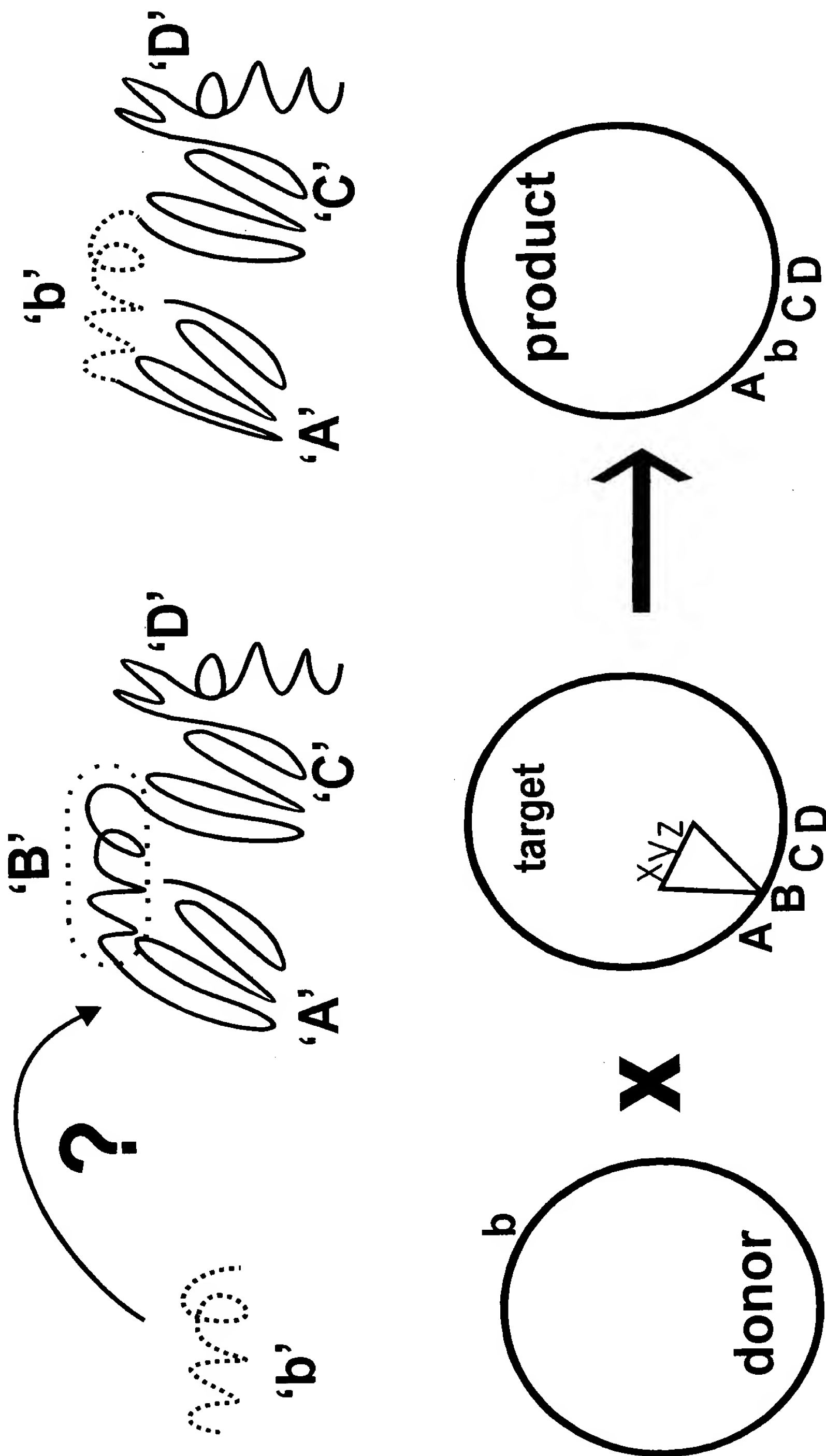
Fragments of family member #2 in donor vector library.



Fragments of family member #3 in donor vector library.

FIG. 7

8/27

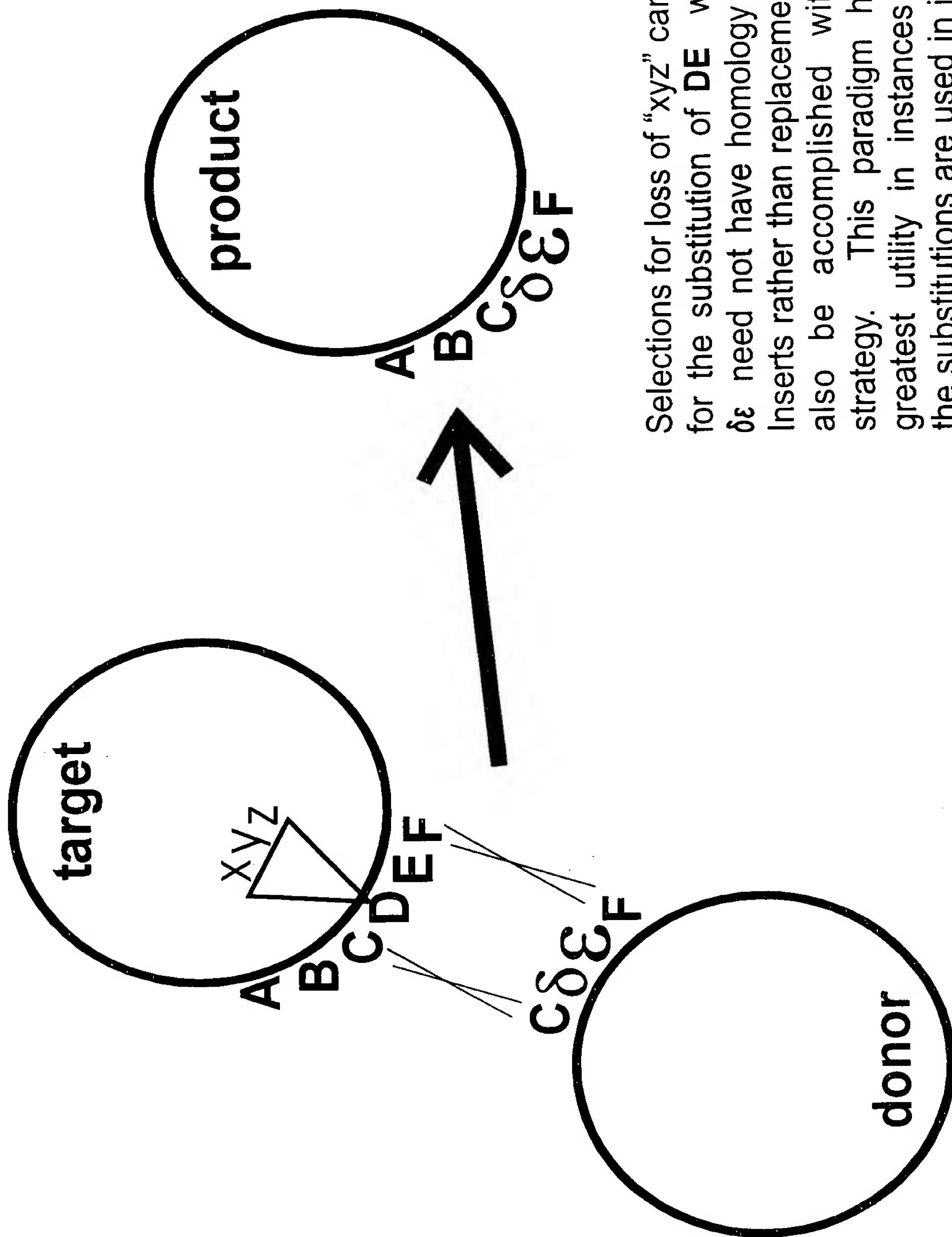


An assay of product function will assess 'b's ability to substitute for 'B'. This paradigm executed with many target proteins and donor domains (and motifs) will identify a library of protein building blocks.

FIG.8



9/27



Selections for loss of "xyz" can select for the substitution of **DE** with  $\delta\epsilon$ .  $\delta\epsilon$  need not have homology to DE. Inserts rather than replacements can also be accomplished with this strategy. This paradigm has the greatest utility in instances where the substitutions are used in iterative combinatorial processes.

FIG.9

10/27

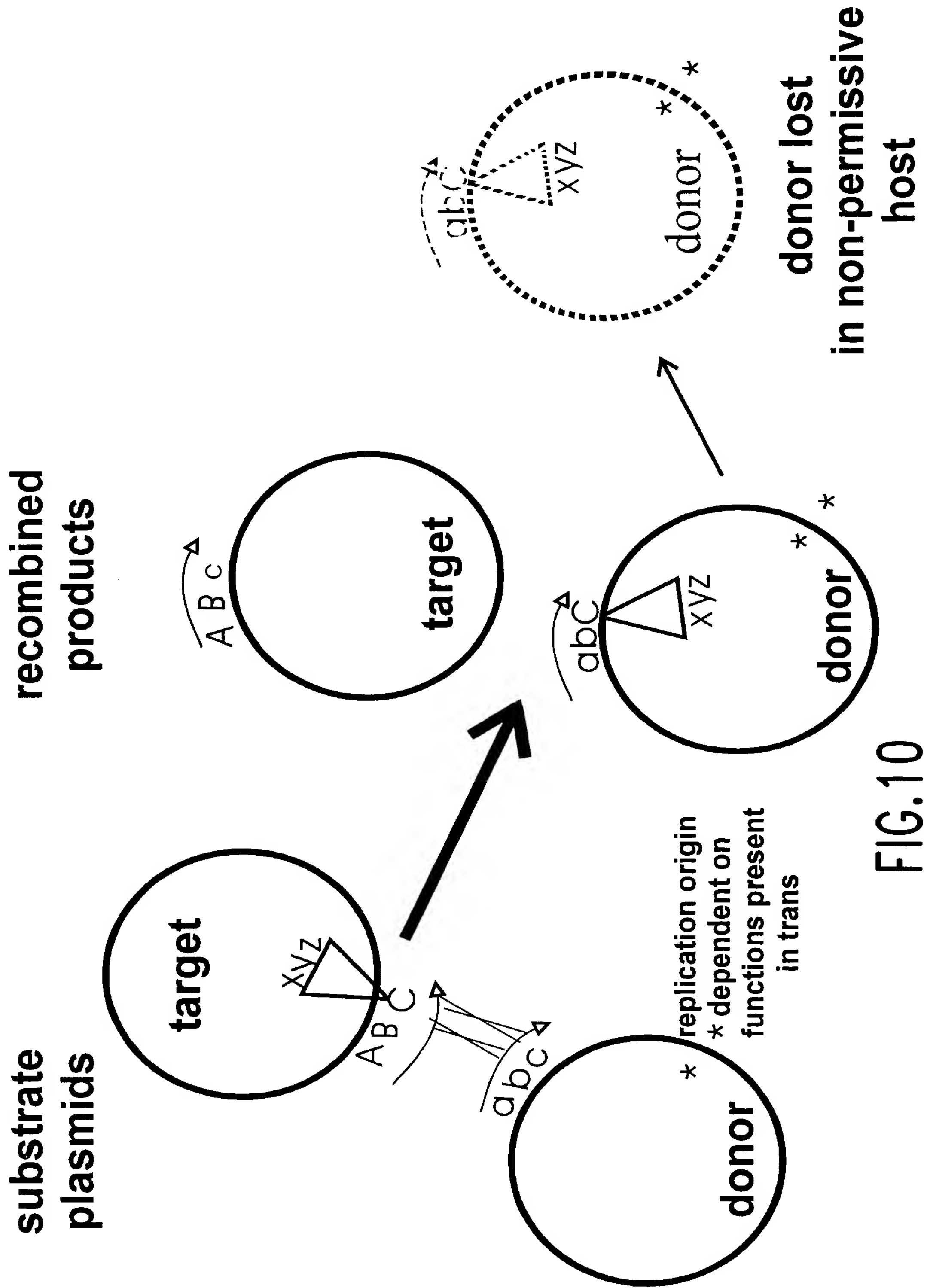


FIG.10

11/27

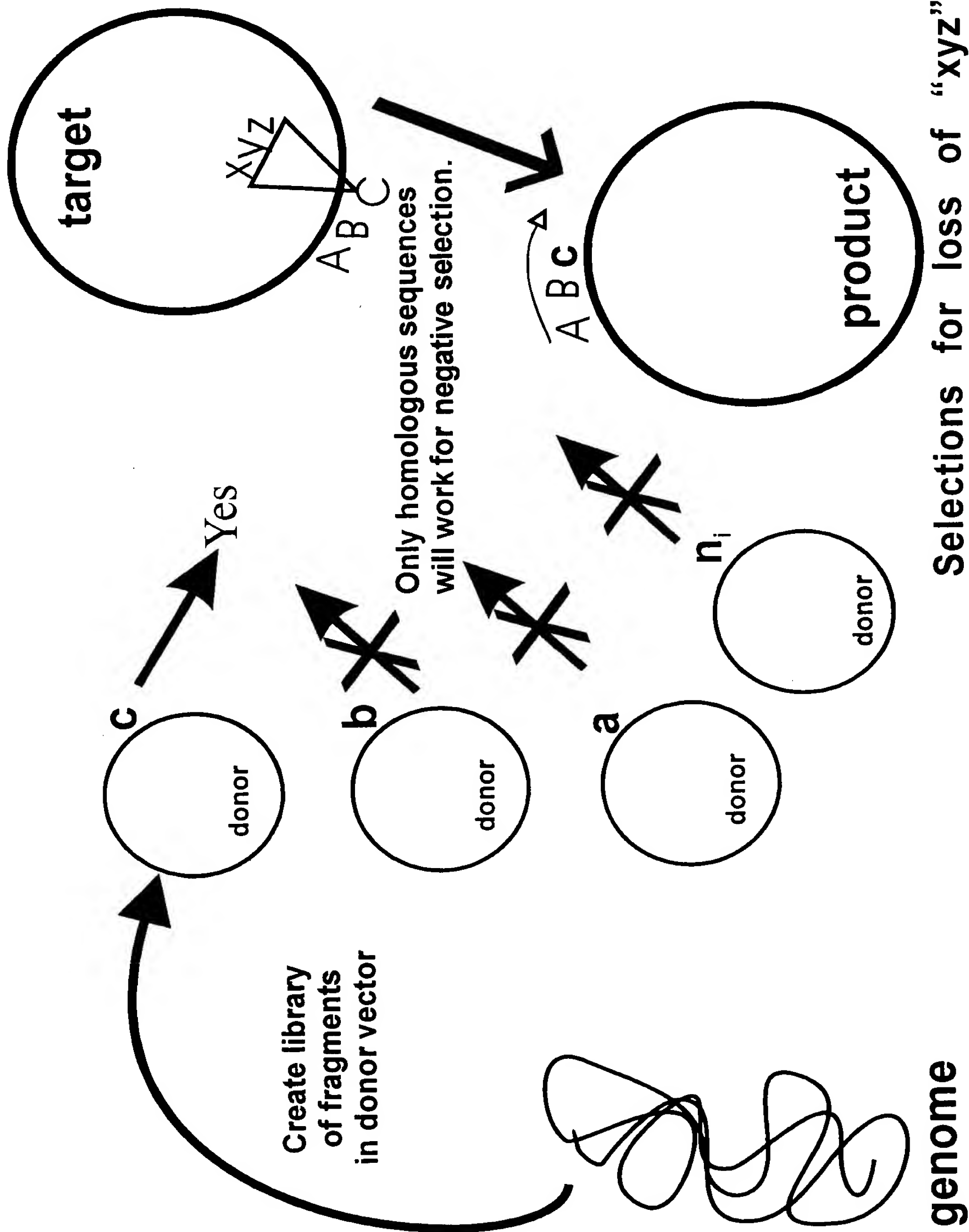


FIG. 11

Selections for loss of "xyz" can select homologous sequences from a complex a donor cloned library.

12/27

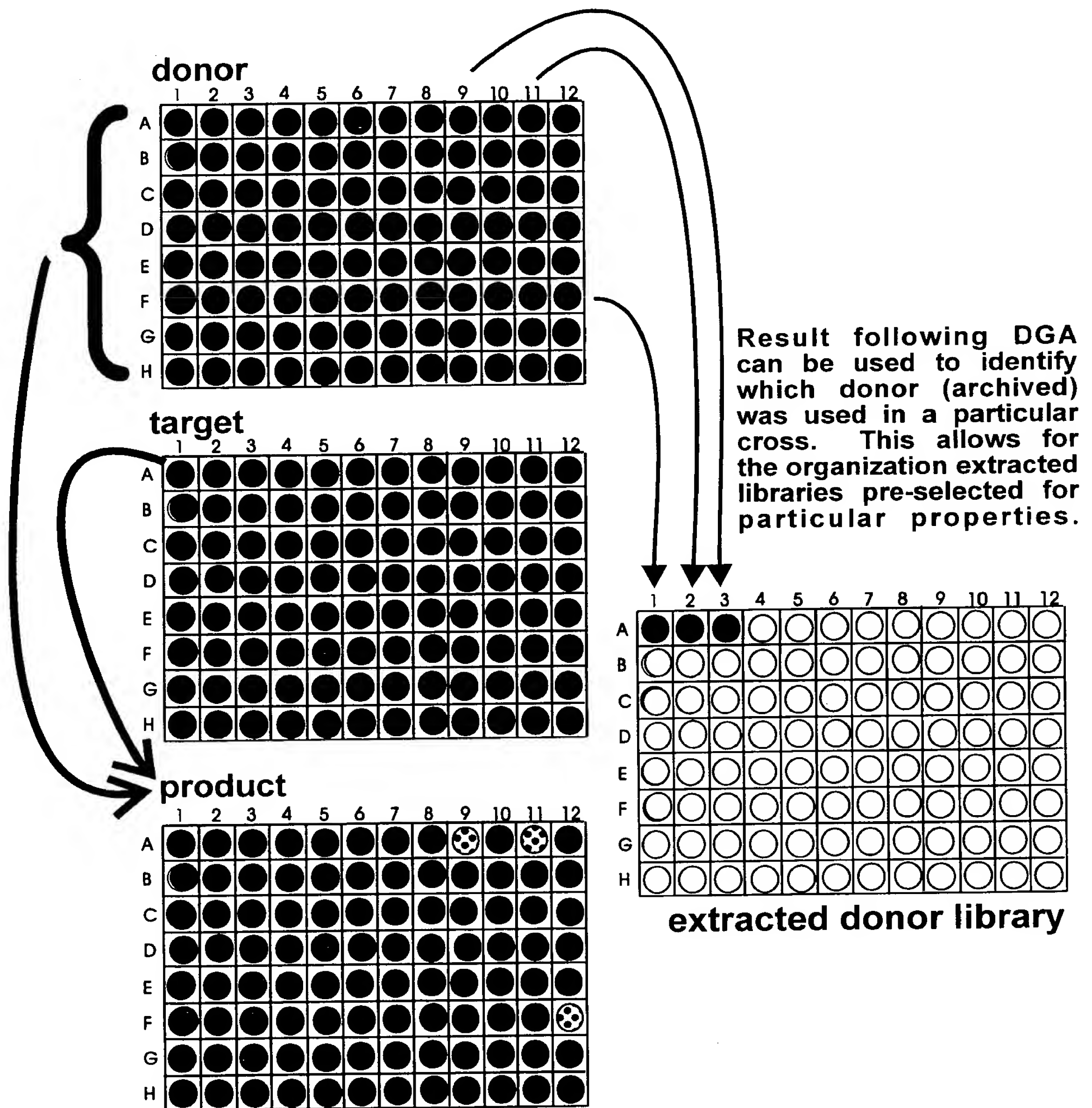
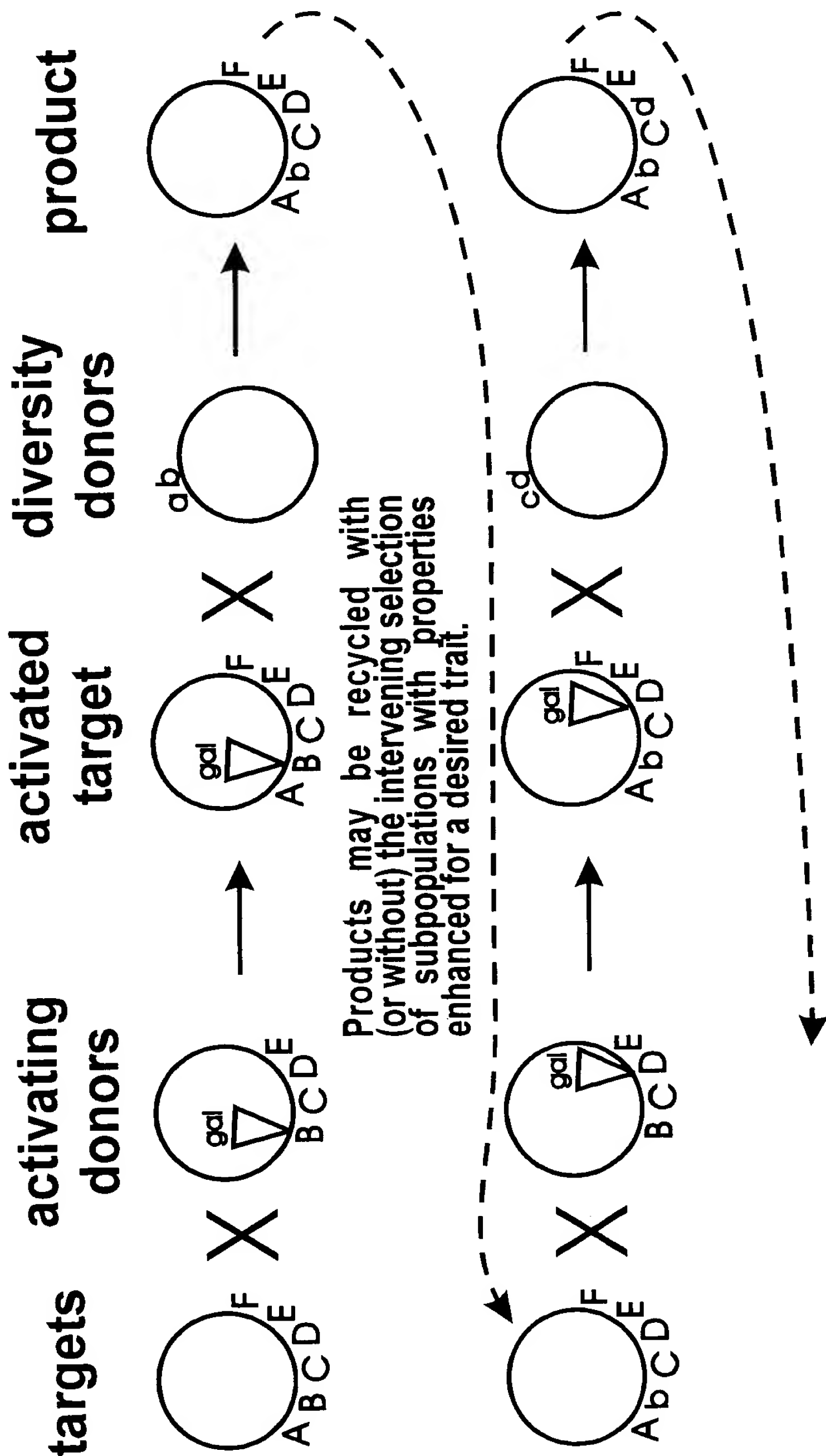


FIG. 12

# DGA Activation

# DGA Diversity



With the product recycled via a DGA activation step iterative cycles are possible. Iterative cycles will allow huge numbers of combinatorial arrangements to be produced from a relatively small defined set of substrates in a systematic process.

FIG. 13

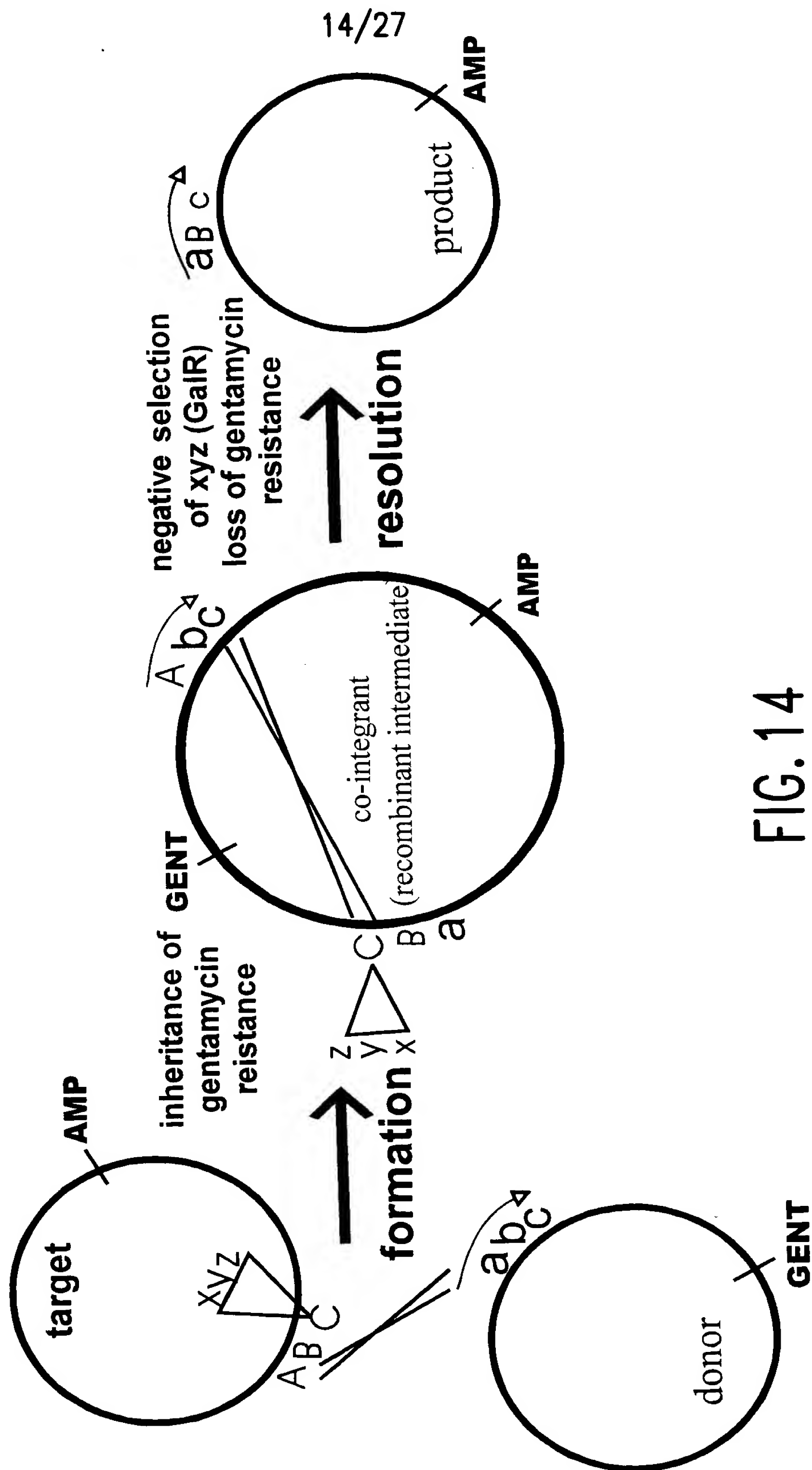


FIG. 14

15/27

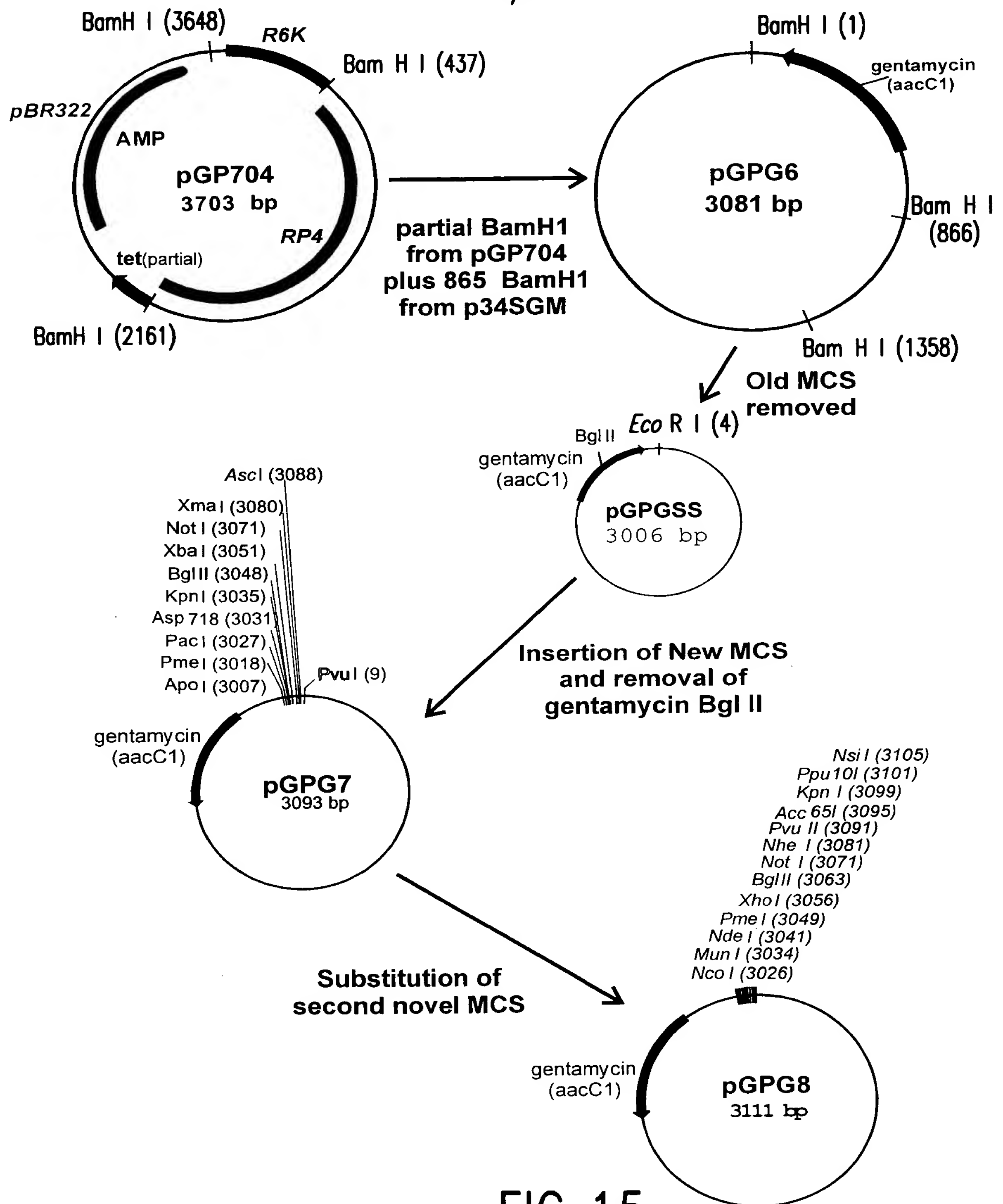


FIG. 15

16/27

+1 I K A D K V Q A Q G F K G A N V K V A V L  
1 ATTAAGCG GACAAAGTG CAGGCTCAA GGCTTTAAG GGAGCGAAT GTAAAAGTA GCCGTCCTG  
  
+1 D T G I Q A S H P D L N V V G G A S F V A  
64 GATACAGGA ATCCAAGCT TCTCATCCG GACTTGAAC GTAGTCGGC GGAGCAAGC TTTGTGGCT  
  
+1 G E A Y N T D G N G H G A H V A G T V A A  
127 GGCGAAGCT TATAACACC GACGGCAAC GGACACGGC GCACATGTT GCCGGTACA GTAGCTGCG  
  
+1 L D N T T G V L G V A P S V S L Y A V K V  
190 CTTGACAAT ACAACGGGT GTATTAGGC GTTGCGCCA AGCGTATCC TTGTACGCG GTTAAAGTA  
  
+1 L N S S G S G S Y S G I V S G I E W A T T  
253 CTGAATTCA AGCGGAAGC GGATCATAC AGCGGCATT GTAAGCGGA ATCGAGTGG GCGACAACA  
  
+1 N G M D V I N M S L G G A S G S T A M K Q  
316 AACGGCATG GATGTTATC AATATGAGC CTTGGGGGA GCATCAGGC TCGACAGCG ATGAAACAG  
  
+1 A V D N A Y A K G V V V V A A A G N S G S  
379 GCAGTCGAC AATGCATAT GCAAAGGG GTTGTCGTT GTAGCTGCA GCAGGGAAC AGCGGATCT

**DNA Sequence 5A20 *B.licheniformis* isolate**

+1 I K A P A L H S Q G Y T G S N V K V A V I  
1 ATTAAGCG CCGGCTCTT CACTCTCAA GGCTACACA GGTTCTAAC GTAAAAGTA GCCGTAATT  
  
+1 D S G I D S S H P D L N V R G G A S F V P  
64 GACAGCGGA ATTGACTCT TCTCATCCT GACTTGAAC GTCAGAGGC GGAGCAAGC TTCGTACCT  
  
+1 S E T N P Y Q D G S S H G T H V A G T V A  
127 TCTGAAACA AACCCATAC CAAGATGGC AGTTCTCAC GGCACACAT GTAGCCGGT ACGGTTGCC  
  
+1 A L N N S I G V L G V A P N A S L Y A V K  
190 GCACTTAAT AACTCAATC GGTGTTTTG GGCGTAGCG CCAAACGCA TCGTTATAT GCAGTAAAA  
  
+1 V L D S T G N G Q Y S W I I N G I E W A I  
253 GTTCTTGAT TCAACAGGA AACGGCCAA TACAGCTGG ATTATTAAC GGCATTGAG TGGGCCATT  
  
+1 S N K M D V I N M S L G G P S G S T A L K  
316 TCCAACAAA ATGGACGTG ATTAACATG AGCCTTGGC GGACCTTCT GGTTCTACA GCTTTGAAA  
+1 S V V D R A V A S G I V V V A A A G N E G  
379 TCAGTCGTT GATAGAGCC GTAGCCAGC GGTATCGTC GTTGTTGCT GCAGCCGGA AATGAAGGC  
  
+1 T S G S S S T I G Y P A K Y P S T I A V G  
442 ACTTCCGGA AGCTCAAGC ACAATCGGC TATCCTGCA AAATATCCT TCTACCATT GCGGTAGGT  
  
+1 A V N S S N Q R G S F S S V G P E L E V M  
505 GCGGTAAAC AGCAGCAAC CAAAGAGGT TCATTCTCA AGCGTAGGT CCTGAGCTT GAAGTCATG  
  
+1 A P G  
568 GCTCCTGGC

**DNA Sequence 3A13 *B.subtilis* isolate**  
**FIG. 16**



17/27

```

      M M R K K S F W L G M L T A L M L V F T M
1  ATGATGAGG AAAAAGAGT TTTTGGCTT GGGATGCTG ACGGCCTTA ATGCTCGTG TTCACGATG

      A F S D S A S A A Q P A K N V E K D Y I V
64 GCCTTCAGC GATTCCGCG TCTGCTGCT CAGCCGGCG AAAAATGTT GAAAAGGAT TATATTGTC

      G F K S G V K T A S V K K D I I K E S G G
127 GGATTTAAG TCGGGAGTG AAAACCGCA TCCGTCAAA AAGGACATC ATCAAAGAG AGCGGCGGA

      K V D K Q F R I I N A A K A K L D K E A L
190 AAAGTGGAC AAGCAGTTT AGAATCATC AACGCGGCA AAAGCGAAG CTAGACAAA GAAGCGCTT

      E E V K N D P D V A Y V E E D H V A H A L
253 GAGGAAGTC AAAAATGAT CCGGATGTC GCTTATGTG GAAGAGGAT CACGTAGCT CATGCTTTG

      A Q T V P Y G I P L I K A D K V Q A Q G Y
316 GCGCAAACC GTTCCTTAC GGCATTCTT CTCATTAAA GCGGACAAA GTGCAGGCT CAAGGCTAC

      K G A N V K V A V L D T G I Q A S H P D L
379 AAGGGAGCG AACGTAAAA GTCGCCGTC CTGGATACA GGAATCCAA GCTTCTCAT CCGGACTTG

      N V V G G A S F V A G E A Y N T D G N G H
442 AACGTAGTC GGCGGAGCA AGCTTCGTA GCTGGCGAA GCTTATAAC ACCGACGGC AACGGACAC

      G T H V A G T V A A L D N T T G V L G V A
505 GGCACGCAT GTTGCCGGT ACAGTAGCT GCGCTTGAC AATACAACG GGTGTATTA GCGTTGCG

      P N V S L Y A V K V L N S S G S G S Y S G
568 CCGAACGTA TCCTTGTAC GCGGTTAAA GTGCTGAAT TCAAGCGGA AGCGGATCT TACAGCGGC

      I V S G I E W A T T N G M D V I N M S L G
631 ATTGTAAGC GGAATCGAG TGGGCGACG ACAAACGGC ATGGATGTT ATCAACATG AGCCTTGGA

      G P S G S T A M K Q A V D N A Y A R G V V
694 GGACCATCA GGCTCAACA GCGATGAAA CAGGCGGTT GACAATGCA TATGCAAGA GGGGTTGTC

      V V A A A G N S G S S G N T N T I G Y P A
757 GTTGTGGCG GCTGCTGGG AACAGCGGA TCTTCAGGA AACACGAAT ACAATCGGC TATCCTGCG

      K Y D S V I A V G A V D P N S N R A S F S
820 AAATACGAC TCTGTCATC GCAGTTGGC GCGGTAGAC CCTAACAGC AACAGAGCT TCATTTTCC

      S V G A E L E V M A P G A G V Y S T Y P T
883 AGCGTCGGA GCAGAGCTT GAAGTCATG GCTCCTGGC GCAGGCGTG TACAGCACT TACCCAACC

      S T Y A T L N G T S M A S P H V A G A A A
946 AGCACTTAT GCAACATTG AACGGAACG TCAATGGCT TCTCCTCAT GTAGCGGGA GCAGCAGCT

      L I L S K H P N L S A S Q V R N R L S S T
1009 TTGATCTTG TCAAAACAT CCGAACCTT TCAGCTTCA CAAGTCCGC AACCGTCTC TCCAGTACG

      A T Y L G S S F Y Y G K G L I N V E A A A
1072 GCGACTTAT TTGGGAAGC TCCTTCTAC TATGGAAA GGTCTGATC AATGTGCGA GCTGCCGCT

      Q *
1135 CAATAA

```

**DNA Sequence 5A36 *B.licheniformis* complete**  
**FIG. 17A**

18/27

V R S K K L W I S L L F A L T L I F T M A  
1 GTGAGAAGC AAAAAATTG TGGATCAGC TTGTTGTTT GCGTTAACG TTAATCTTT ACGATGGCG

F S N M S A Q A A G K S S T E K K Y I V G  
64 TTCAGCAAC ATGTCTGCG CAGGCTGCC GGAAAAAGC AGTACAGAA AAGAAATAC ATTGTCGGA

F K Q T M S A M S S A K K K D V I S E K G  
127 TTTAAACAG ACAATGAGT GCCATGAGT TCCGCCAAG AAAAAGGAT GTTATTTCT GAAAAAGGC

G K V Q K Q F K Y V N A A A A T L D E K A  
190 GGAAAGGTT CAAAAGCAA TTTAAGTAT GTTAACGCG GCCGCAGCA ACATTGGAT GAAAAAGCT

V K E L K K D P S V A Y V E E D H I A H E  
253 GTAAAAGAA TTGAAAAAA GATCCGAGC GTTGCATAT GTGGAAGAA GATCATATT GCACATGAA

Y A Q S V P Y G I S Q I K A P A L H S Q G  
316 TATGCGCAA TCTGTTCTT TATGGCATT TCTCAAATT AAAGCGCCG GCTCTTCAC TCTCAAGGC

Y T G S N V K V A V I D S G I D S S H P D  
379 TACACAGGC TCTAACGTA AAAGTAGCT GTTATCGAC AGCGGAATT GACTCTTCT CATCCTGAC

L N V R G G A S F V P S E T N P Y Q D G S  
442 TTAAACGTC AGAGGCGGA GCAAGCTTC GTACCTTCT GAAACAAAC CCATACCAG GACGGCAGT

S H G T H V A G T I A A L N N S I G V L G  
505 TCTCACGGT ACGCATGTA GCCGGTACG ATTGCCGCT CTTAATAAC TCAATCGGT GTTCTGGGC

V A P S A S L Y A V K V L D S T G S G Q Y  
568 GTAGCGCCA AGCGCATCA TTATATGCA GTAAAAGTG CTTGATTCA ACAGGAAGC GGCCAATAT

S W I I N G I E W A I S N N M D V I N M S  
631 AGCTGGATT ATTAACGGC ATTGAGTGG GCCATTTC AACAATATG GATGTTATC AACATGAGC

L G G P T G S T A L K T V V D K A V S S G  
694 CTTGGCGGA CCTACTGGT TCTACAGCG CTGAAAACA GTCGTTGAC AAAGCCGTT TCCAGCGGT

I V V A A A A G N E G S S G S T S T V G Y  
757 ATCGTCGTT GCTGCCGCA GCCGGAAC GAAGTTCA TCCGGAAGC ACAAGCACA GTCGGCTAC

P A K Y P S T I A V G A V N S S N Q R A S  
820 CCTGCAAAA TATCCTTCT ACTATTGCA GTAGGTGCG GTAAACAGC AGCAACCAA AGAGCTTCA

F S S A G S E L D V M A P G V S I Q S T L  
883 TTCTCCAGC GCAGGTTCT GAGCTTGAT GTGATGGCT CCTGGCGTG TCCATCCAA AGCACACTT

P G G T Y G A Y N G T S M A T P H V A R A  
946 CCTGGAGGC ACTTACGGC GCTTATAAC GGAACGTCC ATGGCGACT CCTCACGTT GCCCGAGCA

A A L I L S K H P T W T N A Q V R D R L E  
1009 GCAGCGTTA ATTCTTTCT AAGCACCCG ACTTGGACA AACGCGCAA GTCCGTGAT CGTTTAGAA

S T A T Y L G N S F Y Y G K G L I N V Q A  
1072 AGCACTGCA ACATATCTT GGAAACTCT TTCTACTAT GGAAAAGGG TTAATCAAC GTACAAGCA

A A Q \*  
1135 GCTGCACAA TAA

DNA Sequence 3A1 *B.subtilis* complete  
FIG. 17B

19/27

**Spectinomycin  
Resistance (aadA)**

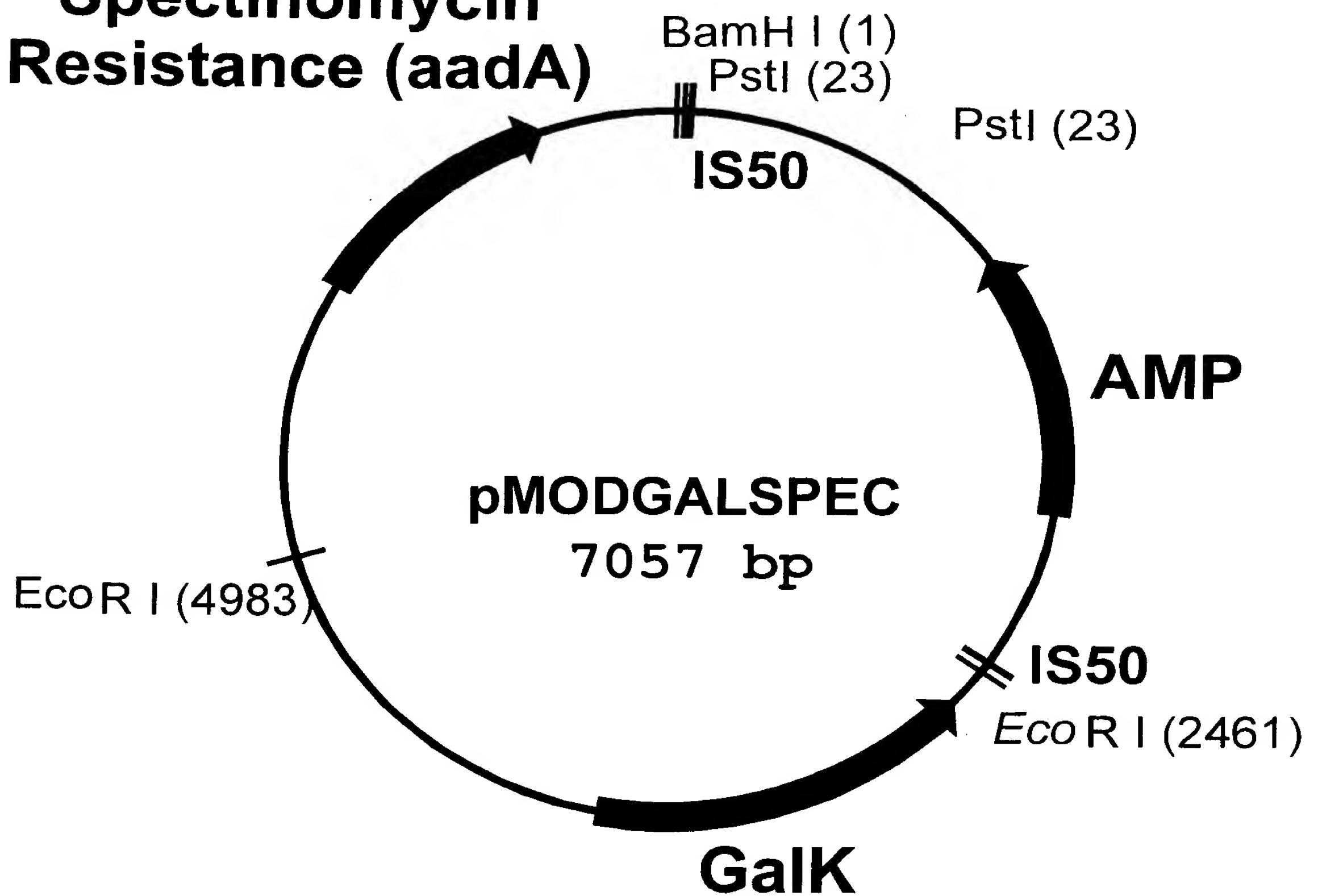


FIG. 18

20/27

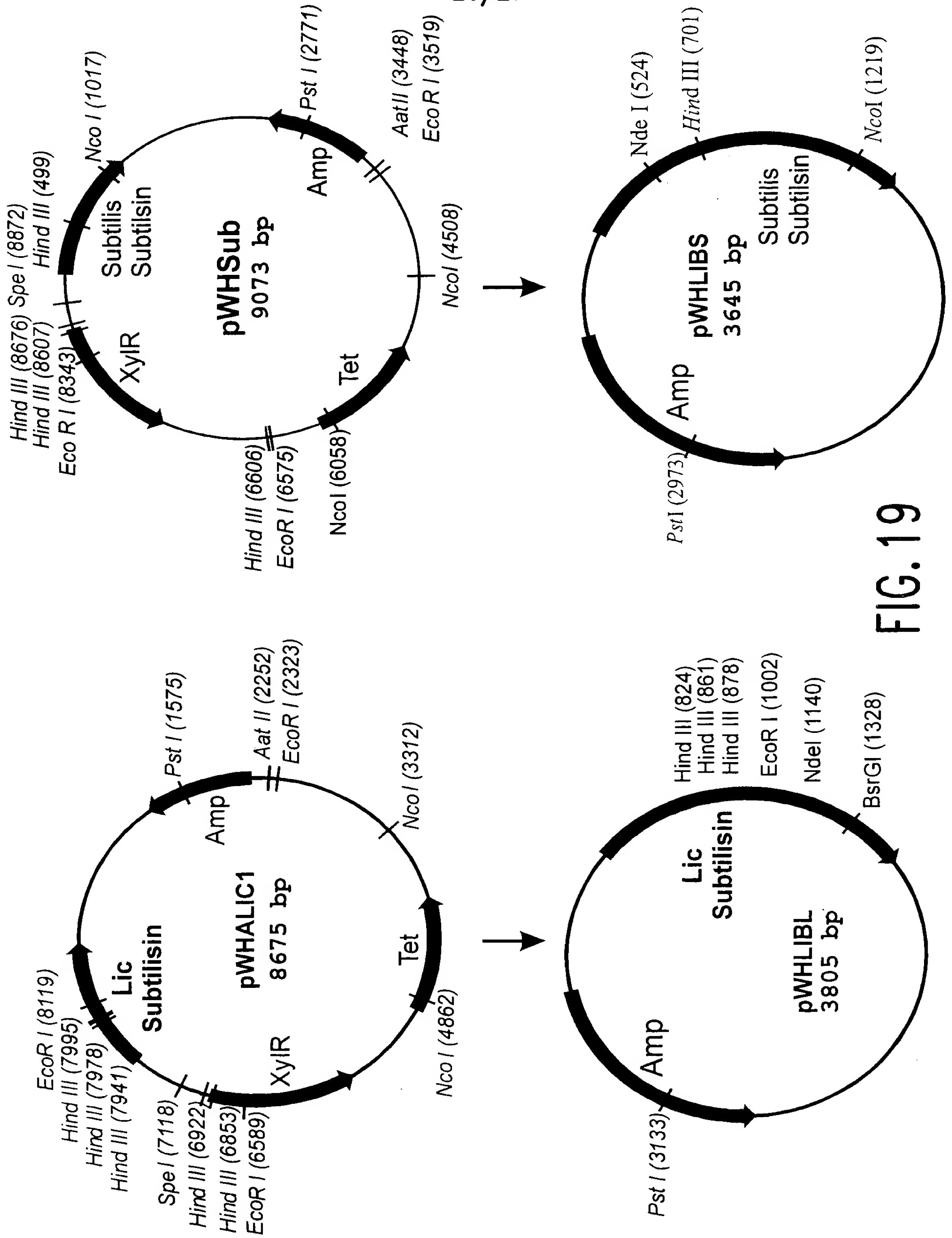


FIG. 19

21/27

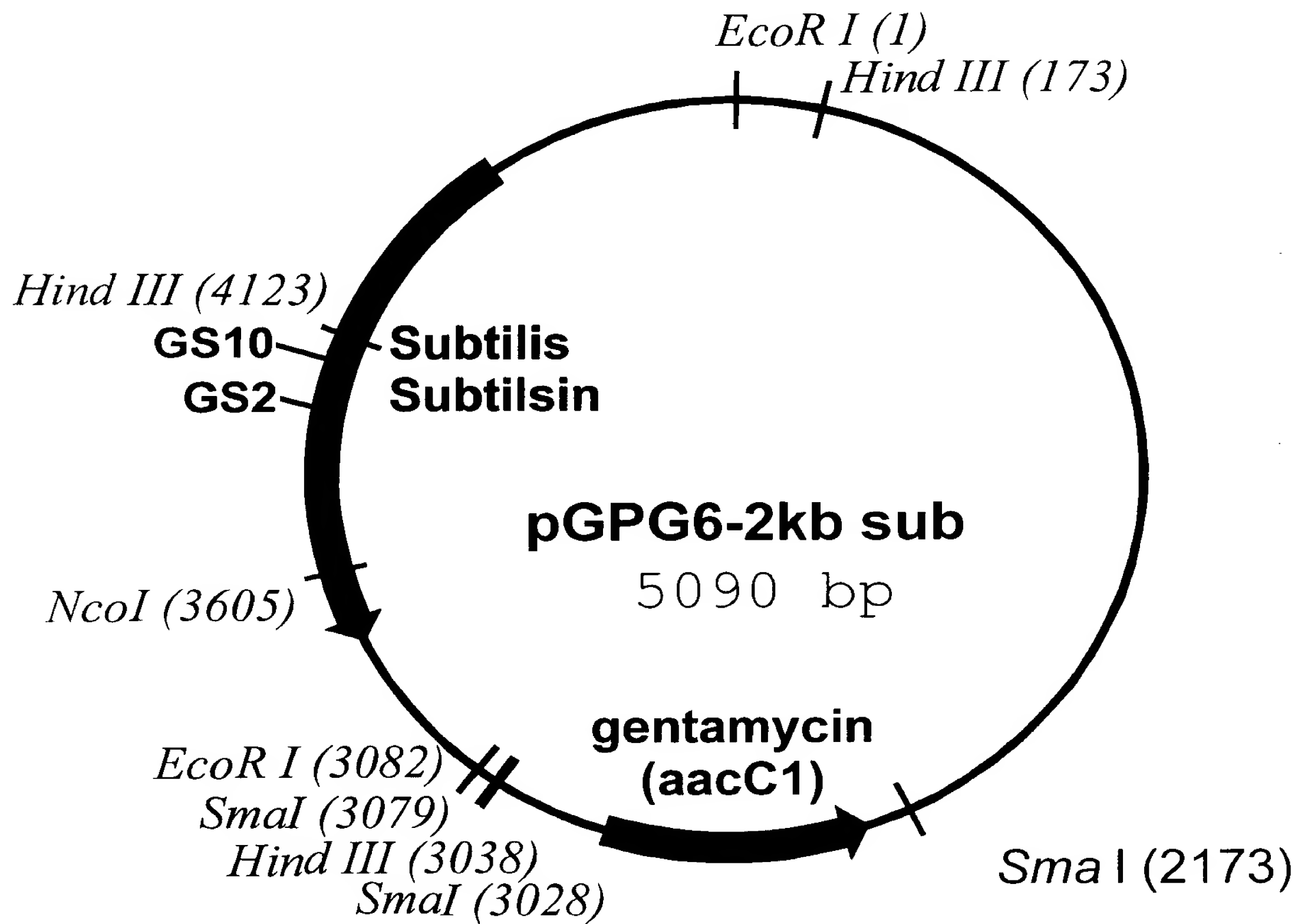
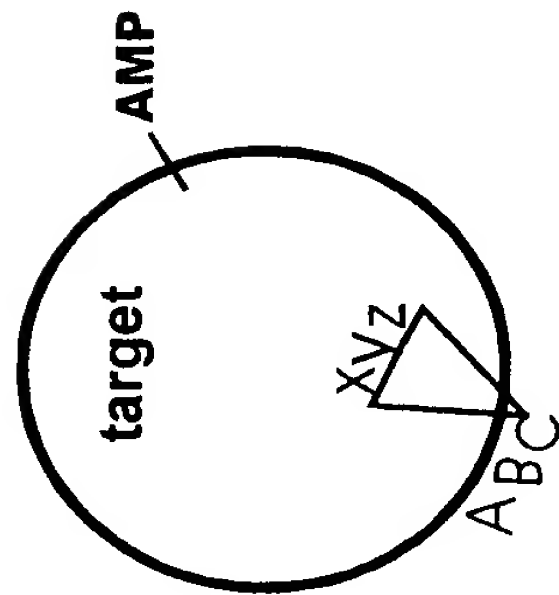


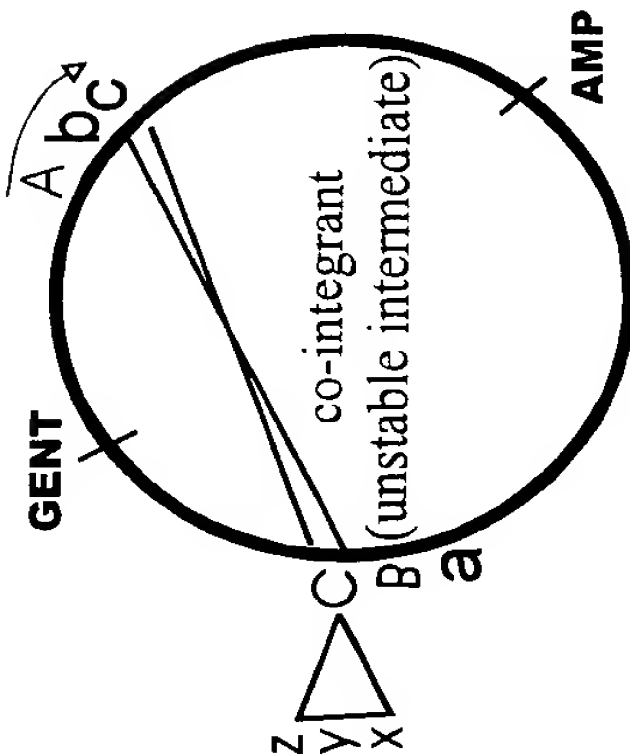
FIG. 20

22/27

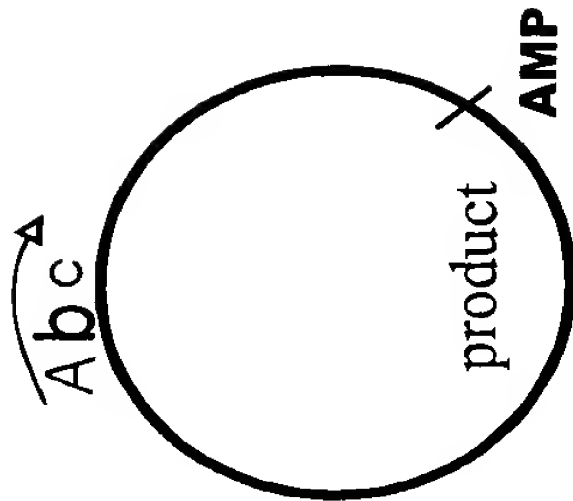
From the co-integrant  
possible molecules include:



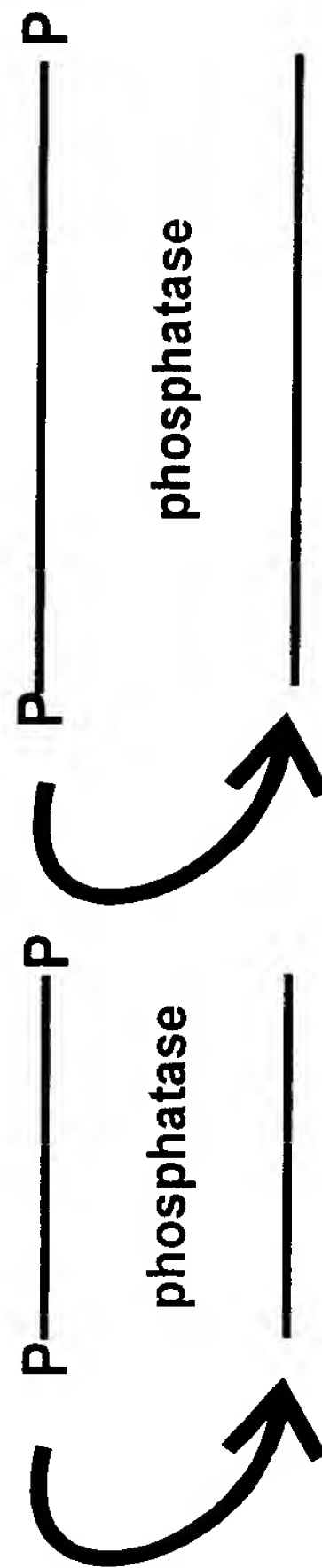
Enzyme digesting  
uniquely in the insert (xyz)  
will cut



Enzyme digesting  
uniquely in the insert (xyz)  
will cut



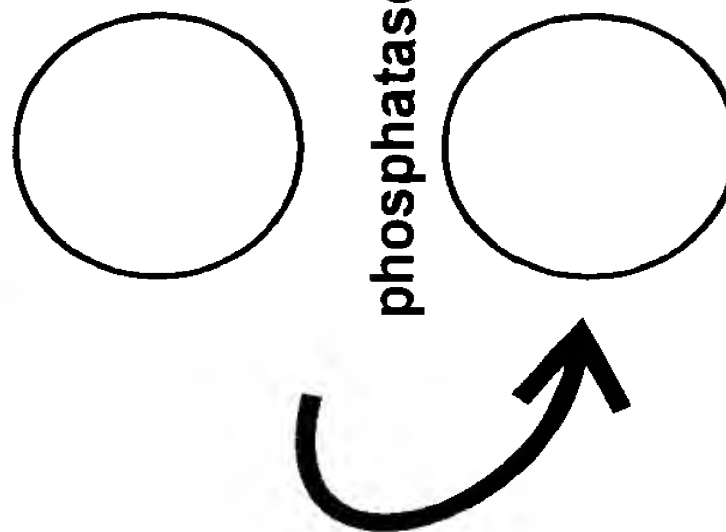
Enzyme digesting  
uniquely in the insert (xyz)  
does not cut



very few transformants

very few transformants

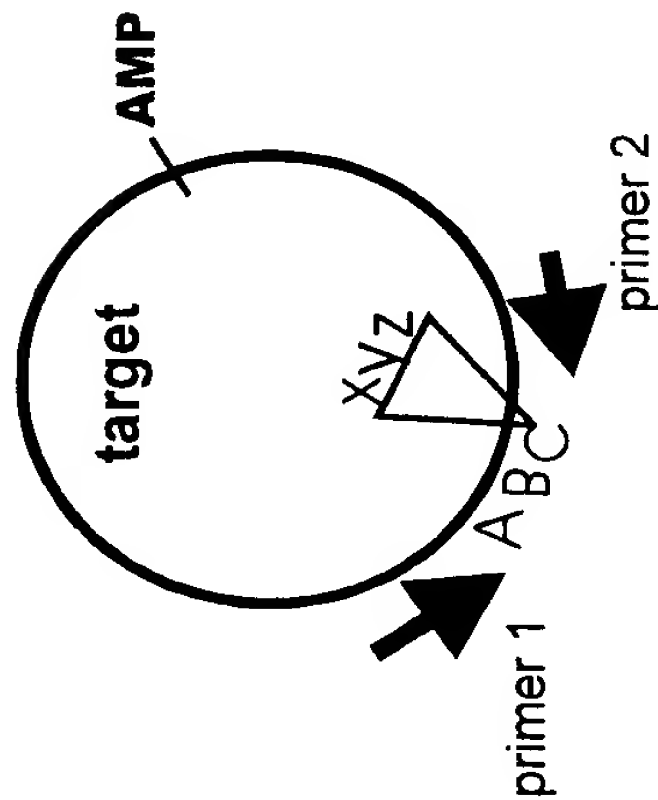
FIG. 21



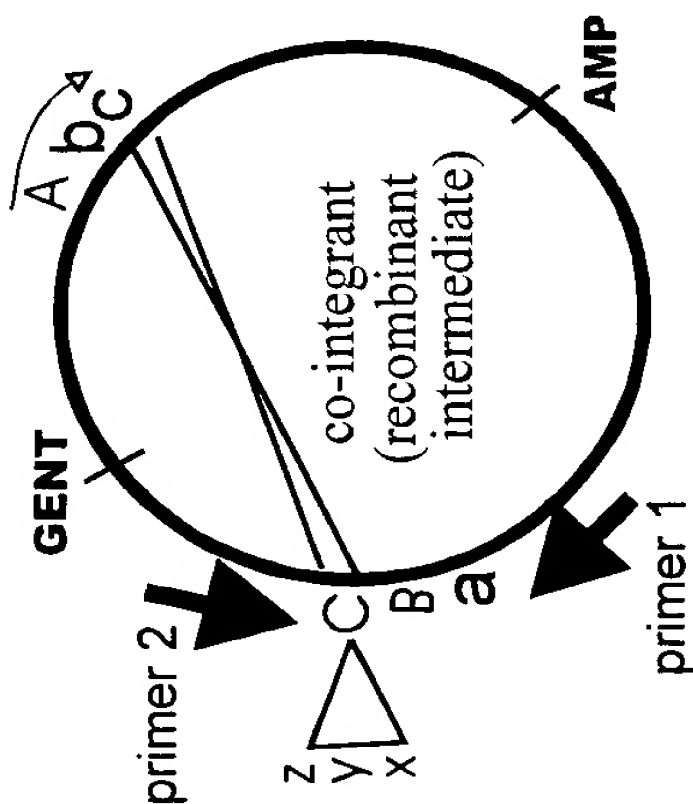
normal number of  
transformants

23/27

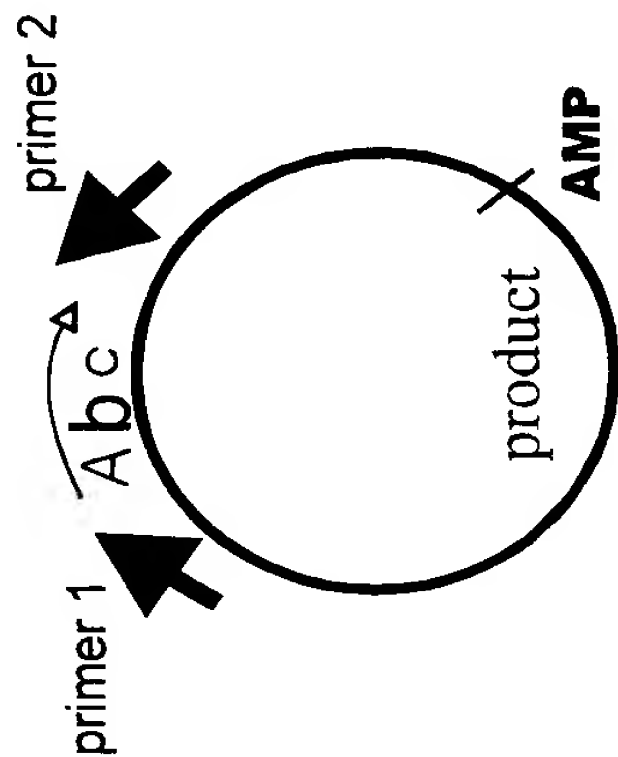
From the co-integrant  
possible molecules include:



**no product  
(too long)**



**no product  
(too long)**



**product**

FIG. 22A

24/27

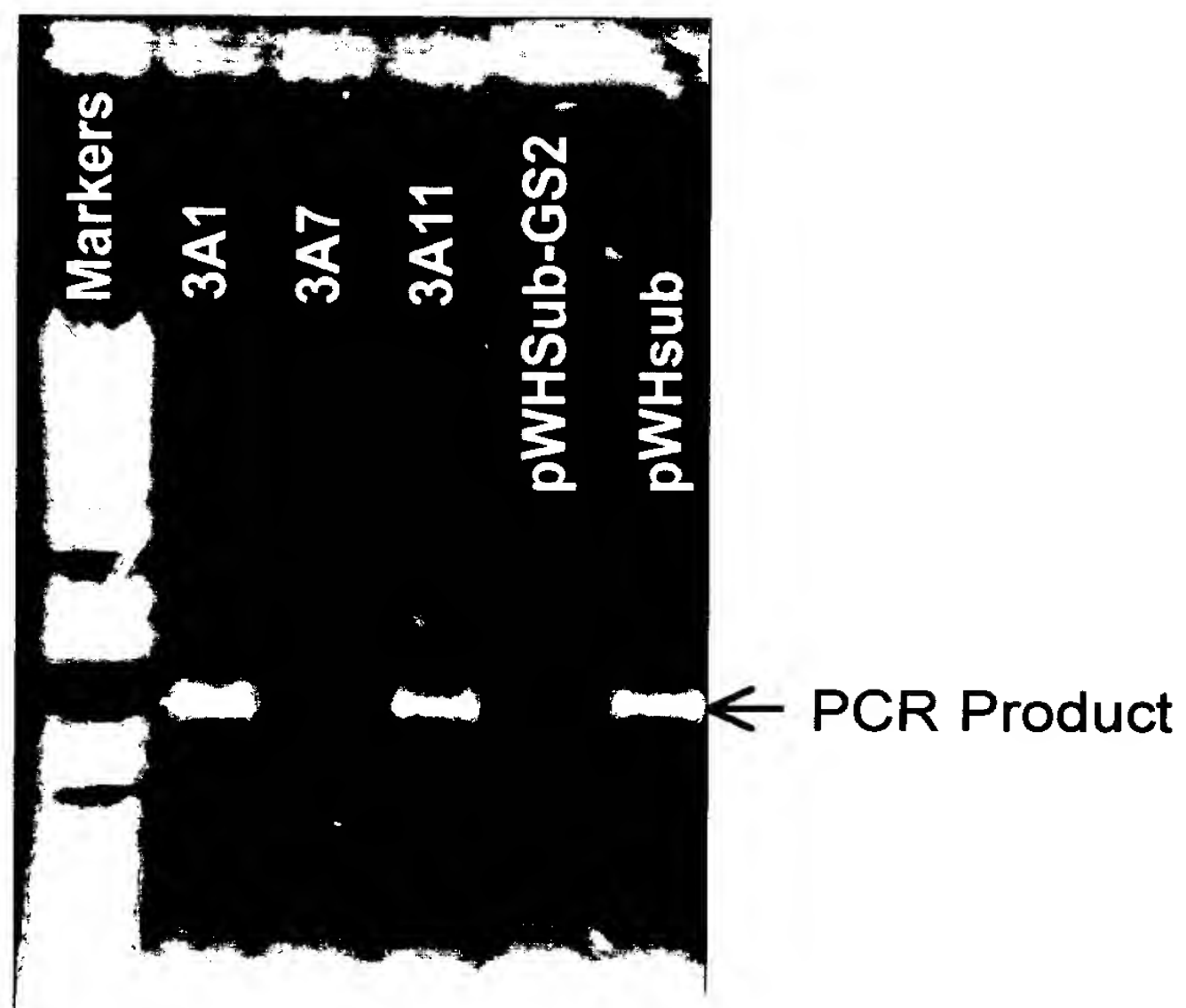


FIG.22B



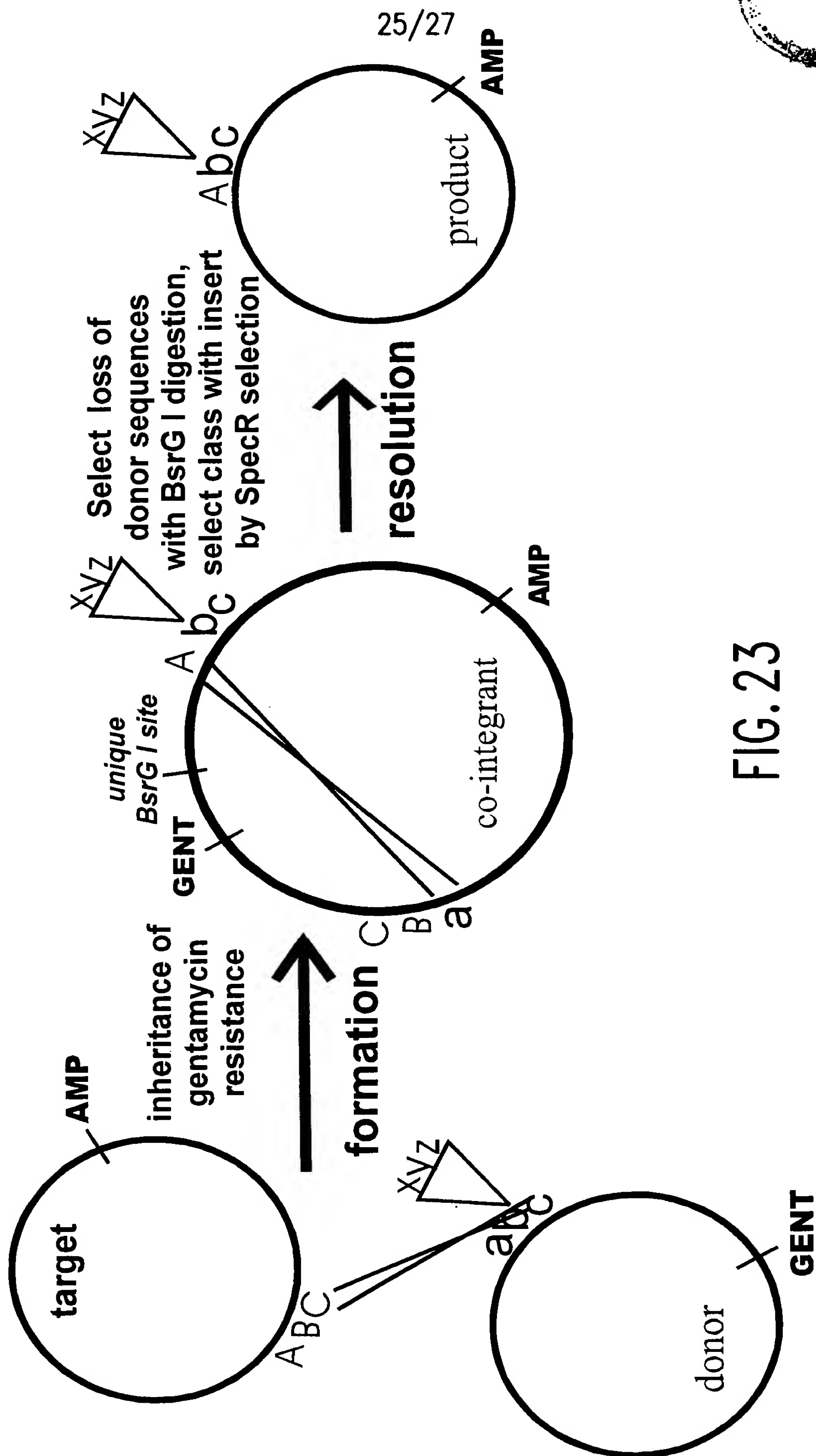
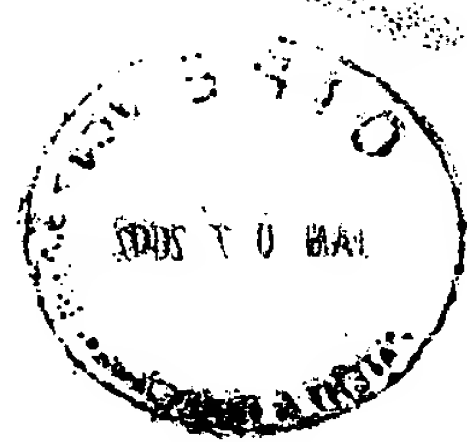


FIG. 23

Oligonucleotide Name	Oligonucleotide SEQ ID NO.	Sequence of oligonucleotide
MCS1F	SEQ ID NO:1	5'-AATCGCGTTAAACTTAATGAAGTACCCATTTTTGGCAGATCTAGACCAAAAA TGGGGCGCGCGCTCCCGGTGGCGGCC-3'
MCS1R	SEQ ID NO:2	5'-AATTGCGCGCCACCCGGGAGCGGCCCCCATTTTTGGTCTAGATCTGCCAAA AATGGTACCTTAATTAAGTTTAAACGCG-3'
BglKF	SEQ ID NO:3	5'-GACTGCGAGATCATAGATATAGATTTCACCTACGGGCTGCTCAACCTGG-3'
BglKR	SEQ ID NO:4	5'-CCAGGTTTGAGCAGCCGGTAGTGAAATCTATATCTATGATCTCGCAGTC-3'
CC_UPPER	SEQ ID NO:5	5'-AATTACCATGGAGCAATTGCATATGGTTAAACAGCTCGAGTAGATCTTGGGGCC GCTGGCTAGCGTCAGCTGGTACCATGCAT -3'
CC_LOWER	SEQ ID NO:6	5'-CGCGTTATGCATGTACCCAGCTGACGCTAGCCAAGCGGCCGAAGATCTACTCGAG CTGTTTAAACCATAATGCAATTGCTCCATGG-3'
internal primer- upper	SEQ ID NO:7	5'-CGCAA(T/A)C(T/C)GTTCCCTA(C/T)GG-3'
internal primer- upper	SEQ ID NO:8	5'-GCCAGGAGCCAT(C/G)AC(A/T)TCAA-3'

**FIG. 24A**

Oligonucleotide Name	Oligonucleotide SEQ ID NO.	Sequence of oligonucleotide
<i>B. lichenformis</i> Subtilisin Forward Primer	SEQ ID NO:9	5'-GGGTACCGCGGCTATTCACTTTCG-3'
<i>B. lichenformis</i> Subtilisin Reverse Primer	SEQ ID NO:10	5'-GCAGATCTCATTTGTTAGATATGTTATTGAGCGGC-3'
<i>B. subtilis</i> Subtilisin Forward Primer	SEQ ID NO:11	5'-AGCGAGATCTCTATTATTGTCAGCTG-3'
<i>B. subtilis</i> Subtilisin Reverse Primer	SEQ ID NO:12	5'-GCGCGGTACCTGATAAAGGAGAGGGTAAAGAG-3'
Galactokinase upper primer	SEQ ID NO:17	5'-GGAAGATCTAGAGGTTTCACCGTCATCACCG-3'
Galactokinase lower primer	SEQ ID NO:18	5'-GGTAGATCTCTTTCGTCGCTTCAAGAAATCCGC-3'

FIG. 24B